153/12.8.70

Psychological Bulletin

478Z 85 A

WAYNE DENNIS, Editor
Brooklyn College

EDWARD GIRDEN, Associate Editor (Book Reviews)
Brooklyn College

ROBERT L. THORNDIKE, Associate Editor (Statistics)
Teachers College, Columbia University

LORRAINE BOUTHILET, Managing Editor

VOLUME 52, 1955

PUBLISHED BIMONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH STREET N.W.
WASHINGTON 6, D. C.

Copyright, 1955, by the American Psychological Association, Inc.

2871

12.8.70
8153

CONTENTS OF VOLUME 52

ARTICLES

The Ability to Judge People: Ronald Taft 1
Psychological Studies of Values: William F. Dukes 24
The Teaching of Psychology: A Survey of Research Since 1942: Robert Birney and Wilbert McKeachie 51
Rotary Pursuit Apparatus: I. Survey of Variables: R. B. Ammons 69
A Note on Gaier and Lee's "Pattern Analysis": Norman H. Livson and Thomas F. Nichols 77
Survey of Research with Psychological Tests in India: W. Leslie Barnette 105
Estimating the Chance Expectancies of Diadic Relationships Within a Group: Renato Tagiuri, Jerome S. Bruner, and Nathan Kogan 122
The Adaptation and Use of Kendall's Coefficient of Concordance (*W*) to Sociometric-Type Rankings: Ben Willerman 132
Complex Tasks for Use in Human Problem-Solving Research: Wilbert S. Ray 134
Processes Affecting Scores on "Understanding of Others" and "Assumed Similarity": Lee J. Cronbach 177
Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores: Paul E. Meehl and Albert Rosen 194
Intra-Individual Response Variability: Donald W. Fiske and Laura Rice 217
Editorial Note 280
Construct Validity in Psychological Tests: Lee J. Cronbach and Paul E. Meehl 281
Psychology in the Arab Near East: E. Terry Prothro and Levon H. Melikian 303
The Role of Motivation in Verbal Learning and Performance: I. E. Farber 311
Learning During Sleep? Charles W. Simon and William H. Emmons 328
Teaching in the Ivory Tower, with Rarely a Step Outside: S. L. Pressey 343
Additions to Psychology Necrology: Japan, 1928–1952: Koji Sato 345
Erratum: Renato Tagiuri, Jerome S. Bruner, and Nathan Kogan 376
Content-Analysis Studies of Psychotherapy: Frank Auld, Jr., and Edward J. Murray. 377
Employee Attitudes and Employee Performance: Arthur H. Brayfield and Walter H. Crockett 396
The "Post-Mortem" Testing of Experimental Comparisons: Richard B. McHugh and Douglas S. Ellis 425
The Infantile Disorders of Hospitalism and Anaclitic Depression: Samuel R. Pinneau. 429
Reply to Dr. Pinneau: René A. Spitz 453
Reply to Dr. Spitz: Samuel R. Pinneau 459
Sound-Precipitated Convulsions: 1945 to 1954: William Bevan 473
Some Perspectives on "The Attenuation Paradox in Test Theory": Frederic M. Lord . . 505

BOOK REVIEWS

Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner's Personality Through Perception: Leo Postman 79
Christie and Jahoda's Studies in the Scope and Method of "The Authoritarian Personality": Roy M. Hamlin 83

Gray's Psychology Applied to Human Affairs: W. J. E. CRISSY 85
Carroll's The Study of Language: IRA J. HIRSH 87
Arnold and Gasson's The Human Person: An Approach to an Integral Theory of Personality: BERT R. SAPPENFIELD 88
Harms' Essentials of Abnormal Child Psychology: MORRIS KRUGMAN 90
Segal's Le Mécanisme de la Vision des Couleurs: MICHAEL J. ZIGLER 91
Perry's Realms of Value: MELVIN H. MARX 93
Brown's Current Theory and Research in Motivation: A Symposium: BERNARD SAPER. 94
McClelland, Atkinson, Clark, and Lowell's The Achievement Motive: ROBERT W. WHITE 95
Anastasi's Psychological Testing: J. P. GUILFORD 97
Fink's Mind and Performance: A Comparative Study of Learning in Mammals, Birds, and Reptiles: W. N. KELLOGG 98
Bellows and Estep's Employment Psychology: The Interview: WILLIAM MCGEHEE 98
Fruchter's Introduction to Factor Analysis: LEONARD S. KOGAN 99
Ausubel's Theory and Problems of Adolescent Development: BARNEY KATZ 100
Bross' Design for Decision: DAVID A. GRANT 101
Recent Books in Child Psychology: BOYD R. MCCANDLESS 150
Maslow's Motivation and Personality; Jones' Nebraska Symposium on Motivation, 1954: DAVID C. MCCLELLAND 159
Estes' Modern Learning Theory. A Critical Analysis of Five Examples: ERNEST R. HILGARD 161
Tinbergen's The Herring Gull's World: FRANK A. BEACH 163
Gillin's For a Science of Social Man: W. J. MCKEACHIE 165
Wolberg's The Technique of Psychotherapy: JULIAN B. ROTTER 167
Sarason's The Clinical Interaction: With Special Reference to the Rorschach: LEONARD D. GOODSTEIN 168
Gannon's Psychology: The Unity of Human Behavior: HENRYK MISIAK 169
Braatøy's Fundamentals of Psychoanalytic Technique: ROY M. HAMLIN 171
Ginzberg, Herma, and Ginsburg's Psychiatry and Military Manpower Policy: LEONARD CARMICHAEL 172
Bray's Issues in the Study of Talent: LEONARD CARMICHAEL 172
Sappenfield's Personality Dynamics: RALPH D. NORMAN 173
Ingham and Love's The Process of Psychotherapy: SEYMOUR SARASON 174
Some Recent Books on Counseling and Adjustment: EDWARD JOSEPH SHOBEN, JR. 251
Carmichael's Manual of Child Psychology: ROGER G. BARKER 263
Thrall, Coombs, and Davis' Decision Processes: LEE J. CRONBACH 267
Moses' The Voice of Neurosis: ROY M. HAMLIN 268
Nuttin's Tâche Réussite et Échec: Théorie de la Conduite Humaine: MICHAEL J. ZIGLER. 269
Kornhauser, Dubin, and Ross' Industrial Conflict: JEROME H. ELY 270
Schafer's Psychoanalytic Interpretation in Rorschach Testing: ROY M. HAMLIN 271
Taylor's Dynamic and Abnormal Psychology: CHARLES N. COFER 272
Thorpe and Schmuller's Contemporary Theories of Learning: LAWRENCE M. STOLUROW. 273
Gruenberg's The Encyclopedia of Child Care and Guidance: LEONARD S. KOGAN 274
Sonnemann's Existence and Therapy: ALBERT ELLIS 275
Pennington and Berg's An Introduction to Clinical Psychology: NICHOLAS HOBBS 276
King's Psychomotor Aspects of Mental Disease: An Experimental Study: JAMES D. PAGE 277
Thelen's Dynamics of Groups at Work: ALFRED MCCLUNG LEE 278
Handbook of Social Psychology: DANIEL KATZ 346
Hoch and Zubin's Depression: JAMES D. PAGE 354
Savage's The Foundations of Statistics: LEONARD S. KOGAN 354

Roback's Destiny and Motivation in Language: CALVIN S. HALL 355
Cooper and Erickson's Time Distortion in Hypnosis: ANDRÉ M. WEITZENHOFFER 356
McKeon's Interrelations of Cultures: CLYDE KLUCKHOHN 357
Ziskind's Psychophysiologic Medicine: JOSEPH E. BARMACK 357
Murphy and Bachrach's An Outline of Abnormal Psychology: FRED MCKINNEY 358
Hoijer's Language in Culture: ROGER W. BROWN 359
Klopfer, Ainsworth, Klopfer, and Holt's Developments in the Rorschach Technique. Vol. I. Technique and Theory: GEORGE W. ALBEE 360
Katz, Cartwright, Eldersveld, and Lee's Public Opinion and Propaganda: IVAN D. LONDON 361
McMillan, Grant, Fitts, Frick, McCullough, Miller, and Brosin's Current Trends in Information Theory: A. CHAPANIS 361
Zilboorg's The Psychology of the Criminal Act and Punishment: FRED MCKINNEY 362
Wolstein's Transference: Its Meaning and Function in Psychoanalytic Therapy: ROY M. HAMLIN 363
Wever and Lawrence's Physiological Acoustics: J. DONALD HARRIS 363
Hall's A Primer of Freudian Psychology: DAN L. ADLER 366
Fairbairn's An Object-Relations Theory of the Personality: DAN L. ADLER 366
Sheldon, Dupertuis, and McDermott's Atlas of Men: NANCY BAYLEY 367
Henry and Short's Suicide and Homicide: GEORGE SIMPSON 368
Lowenfeld's The Lowenfeld Mosaic Test: MORRIS KRUGMAN 369
Bellak's The Thematic Apperception Test and the Children's Apperception Test in Clinical Use: MAX L. HUTT 370
Comment Concerning Sappenfield's Review of Arnold and Gasson's The Human Person: MAGDA B. ARNOLD 372
Good and Scates's Methods of Research: MALCOLM G. PRESTON 463
Brand's The Study of Personality: BERT R. SAPPENFIELD 465
Kerman's What Is Electroshock Therapy?: JAMES D. PAGE 466
Ryan and Smith's Principles of Industrial Psychology: MILTON L. BLUM 466
Oeser and Hammond's Social Structure and Personality in a City, and Oeser and Emery's Social Structure and Personality in a Rural Community: DAN L. ADLER 468
Burt's The Causes and Treatment of Backwardness: SEYMOUR B. SARASON 469
Kugelmass' The Management of Mental Deficiency in Children: KARL F. HEISER 470
Edwards' Statistical Methods for the Behavioral Sciences: LEONARD S. KOGAN 470
Le Beau's Psycho-chirurgie et Fonctions Mentales: H. J. EYSENCK 471
Some Recent Texts in Educational Psychology: DONALD SNYGG 511
Stone and McNemar's Annual Review of Psychology. Vols. 5 and 6: JAMES DREVER 518
Allport's Theories of Perception and the Concept of Structure: JOHN F. DASHIELL 521
Révész' Thinking and Speaking: A Symposium: FRANK A. PATTIE 523
Special Devices Center, Office of Naval Research's Handbook of Human Engineering Data: ARNOLD M. SMALL 524
McNemar's Psychological Statistics: LEONARD S. KOGAN 525
Piaget's The Construction of Reality in the Child: JOHN E. ANDERSON 526
Rogers and Dymond's Psychotherapy and Personality Change: VICTOR RAIMY 528
Gring's Laboratory Instrumentation in Psychology: FRANK W. FINGER 529
Beck's The Six Schizophrenias: Reaction Patterns in Children and Adults: GEORGE F. J. LEHNER 530
Kendall's Conflict and Mood: RALPH R. CANTER 531
Piéron, Reuchlin, Bize, Bénassy-Chauffard, Pacaud, and Rennes' Traité de Psychologie Appliquée. Livre Troisième. L'Utilization des Aptitudes. Orientation et Sélection Professionnelles: MICHAEL J. ZIGLER 533

Bach's Intensive Group Psychotherapy: NICHOLAS HOBBS 534
Lewis' Deprived Children: SAMUEL R. PINNEAU 536
Arnheim's Art and Visual Perception: HARRY HELSON 537
Meehl's Clinical Versus Statistical Prediction: LEONARD S. KOGAN 539
Rotter's Social Learning and Clinical Psychology: SEYMOUR B. SARASON 540
Hyman's Interviewing in Social Research: LEONARD S. KOGAN 542
Erratum: ROGER G. BARKER 544

MISCELLANEOUS

Books and Monographs Received 102, 175, 374
Index of Authors 545

VOL. 52, NO. 2 MARCH, 1955

Psychological Bulletin

478405-A

SURVEY OF RESEARCH WITH PSYCHOLOGICAL TESTS IN INDIA

W. LESLIE BARNETTE[1]
University of Buffalo

Within the compass of this article, it is impossible to deal with all the aspects of psychological work now going on in India. Largely because of the interests of the writer, the main emphasis in this summary and critique will be on current research in the area of psychological tests although other areas will be treated. Omitted here is the large volume of work in social psychology, most of which stems from the pioneering work of the Unesco Tensions Project in 1950 for which Gardner Murphy was the consultant. Murphy's recent American publication (**22**), summarizing this research, is easily available to United States readers; consequently it will not be summarized here.

The Indian psychological scene is difficult to summarize in a few words. The orientation is still quite British, largely because the older psychologists, who could afford to go abroad for graduate training, either went to London or to Edinburgh. Here they met the statistical and cognitive traditions of Spearman and Thomson. Consequently, one still finds in India much preoccupation with tests of *g* and practically no diagnostic work. Alexander's Passalong Test (**1**) is widely used; Cox's mechanical ability measures (**12**) are also well known and used. The psychological pole is now definitely shifting toward the United States with the result that there is more interest in "applied" problems and projective tests. Three clear psychological lines are now evident: (*a*) pyschological testing for school use—intelligence, achievement, aptitude; (*b*) studies with applied problems—such as industrial fatigue, research on tests for worker selection, social psychology problems (largely via questionnaires) and tensions between social groups and classes; (*c*) clinical, taking the form of a preoccupation with the Rorschach and TAT. Such clinical testing is typically in the hands of poorly trained individuals; "experts" are apt to appear on the scene after one reading of Klopfer and Kelly; there are as many "local adaptations" of the TAT as there are schools where this type of research secures sponsorship. The situation in this respect is such that one of the recommendations of the members of the workshop in educational-vocational guidance, led by the writer in Delhi in the spring of 1953, stated that "projective techniques should be used only by specially trained persons" (**5**, p. 42).

---

[1] The author was a Fulbright Visiting Professor in Psychology and Vocational Guidance at the Central Institute of Education, Delhi, India, for the academic year 1952–53. All "current" comments relate to this period. During the year the writer travelled extensively and managed to visit all the important universities where psychology is taught.

Departments of psychology are still largely under the dominance of philosophy. The situation in this regard is much like that in the United States before World War I. There are two or three universities in India which can now boast of autonomous psychology departments. What few published Indian textbooks there are in psychology (almost all in English, such being the medium of practically all university education) reflect the old partnership with philosophy. Recent texts to some extent shift away from the functional-structural content and are apt to deal with more "applied" problems; still, the old tradition dies hard. McDougall is still very much in evidence. One of the most recent texts (15), published in 1952, still is based on a highly eclectic approach that attempts to straddle both the philosophical as well as psychological guideposts: beginning with a discussion of the mind-body problem, it moves on to the area of the physiological (nervous system, receptors, sensation and perception), attention, memory, imagination, feeling and emotion, instincts (the McDougall treatment), learning (conditioning is discussed but in the light only of Pavlov and Watson), sentiments, mental measurements (mostly IQ), abnormal psychology, and parapsychology.[2] Contrary to other texts, this book is excellently printed and would sell for something like $3.50 in United States currency—a high price for most students and teachers to pay.

The great preoccupation, however, is still with the IQ and mental tests. There is little awareness of the more basic problems and theoretical contributions from America as regards learning and motivation theory. Statistics, in line with the cognitive emphasis from Great Britain, is much developed, particularly factor analysis. Murphy, during his stay in India under the auspices of Unesco, was requested by the Ministry of Education of the Government of India to be the leader of a conference of Indian psychologists who drew up excellent recommendations for the teaching of psychology and for "modernizing" the typical university set-ups. Committees of several types were appointed, ostensibly to continue on the basis of this groundwork. It is disappointing to report that the original enthusiasm generated by these group discussions, out of which some realistic and relatively inexpensive programs were suggested, has as yet resulted in very little practical action.

Psychology training for Indian students is still largely confined to the graduate level. Most undergraduates get only one course in educational psychology and practically nothing else in social science. Heavy doses of philosophy and history are administered instead. The result is that students at the MA level, typically in schools of education (where they get their first taste of psychology), are ill-prepared for any advanced work. They have little idea of what an experiment is, let alone how to design one; no understanding of elementary statistics; little appreciation of the social forces shaping their own country. But they are willing and avid and eager and, more and more, so it seemed to this writer, to be looking to the United States for "practical" answers to tough problems in social science. The United States, from the point of view of training in psychology, is now looked upon fairly universally in India as an "open sesame." Most serious students hope to obtain an American scholarship; in the fall of 1952, the U. S. Educational Foundation in India received approximately 3,000 applications for about 30 Fulbright student grants for graduate study in America and for all areas of graduate training.

---

[2] I expected to find a good deal of interest in this area and in the works of Rhine. Quite the contrary, few Indian psychologists seemed to know much about it. I attributed this to the tardiness of "current" psychological data being absorbed by Indian workers despite the availability of many recent American journals.

Still, there is much good work being done in India although it seemed to this writer to be heavily on the side of IQ testing and tests of *g*. But there is much interest in American developments and earnest attempts are made to adapt foreign materials to the local scene. Some of this has been wisely called into question; more sophisticated workers are beginning to sound the call that tests transposed from Western cultures rarely work out satisfactorily (**25**); Dayal (**13**) has urged more centralized research in the construction of psychological tests in India and has complained about the overlapping or unnecessary duplication in psychometric research.

The start of work in psychological testing was the project by Rice in Lahore (now situated in Pakistan) which was begun in 1922. Rice was a missionary who early saw the need for a measure of intelligence in mission schools and elsewhere. His pioneer work (**27**), still in use for measurement purposes in many parts of Northern India today, was an Indian language (Hindustani and Punjabi) version of the Terman 1916 revision of the Binet scale with the addition of some nonverbal items from Pintner and Paterson. This is a point-scale adaptation standardized on 929 school boys, 5 to 16 years of age from 37 different schools and six caste groups. Later protocols were added so that the final table of point scores and MA equivalents is based on an *N* of 1,388 (**27**, p. 72).[3]

Thirty-five test items, including nine nonverbal ones, comprise the scale. A Brief Scale of 10 items is provided; it is also possible to employ the nonverbal items as a separate scale but only for cases where there is an obvious linguistic difficulty. Early attempts were made to construct an age scale and this early age scale correlated .98 with the final point scale. The Punjabi version is equivalent phrase by phrase to the Hindustani; only 27 new words had to be introduced for the entire scale so that a separate standardization was therefore not involved. Inspection of Rice's statistical tables shows that the mean or median IQ's for ages 5–16 all cluster around 100 fairly well. There is, however, considerably more variability (σ of 25.95, calculated by the writer from Rice's data) than Terman reported. Rice attributes this to chance errors of both measurement and in the CA's (age data in India present far more difficulties than in western countries). Point scores show the expected increase with CA; test-retest data for 111 cases over a two-year interval gave a correlation of .86 (**27**, p. 140). On the average there was a gain of 7.5 IQ points on such retests, increases being greater where less experienced examiners conducted the first tests. Attempts were made to employ outside validity checks: the *r* between test scores and school marks categorized into five groups for 630 cases was .25 (Rice believes that, because of the character of the Indian school organization, this is possibly as significant as the .45 value reported by Terman).

---

[3] In the writer's extensive travels in India, he never came across a full text of this work. Various mimeographed copies of the test instructions only were available and these were being used indiscriminately for all types of subjects in many parts of India. It was not until the writer's return to the United States that a copy of the full text became available so that the statistical data might be examined.

Part III of Rice's work discusses test results with "Depressed Classes" or Untouchables. A good deal of statistical care went into this section. Results are reported for a selected group of 500 (50 from each of 5 caste groups matched for both age and school progress; 50 from each of the caste groups matched only for age). Here four of the five groups tested scored approximately 100; Brahmins, the Hindu intellectual aristocracy, averaged 107 (**27**, p. 173). Unfortunately, it was not possible to match on the triple basis of age, grade, and years of schooling. Rice's general conclusion is that there is no essential inferiority in the Depressed Class group and that the differences found between the other four caste groups were slight.

A somewhat more sophisticated test is the version of the 1937 Terman-Merrill scale by Kamat in Bombay done under the sponsorship of Godfrey Thomson in Great Britain. (**16**). Owing to the language problem in India, this scale may be profitably used only within certain states in and around Bombay. Kamat's work uses the age scale of Binet and is available in two languages: Kannada (*N* of the standardization group=739) and Marathi (*N*=335). The two standardization groups were located in a middle-sized town of 30,000 population in Bombay State; the author states that the sample is "fairly representative of the general population of Indian children" (**16**, p. 45). The MA levels for the test items were assigned by the method advocated by Cyril Burt, i.e., a test passed by 50 per cent of all six and seven year olds is assigned to VII. Inspection of Kamat's tables shows that, in the main, the mean IQ's for the various age groups are very close to 100 but the *SD* for the entire standardization group of 1,074 children is 18.7. The final scale comprises 13 age groups against Terman's 12; 99 individual tests as against Terman's 90. The validity of the scale as a whole was checked against secondary school teacher estimates; the author remarks this was "nearly .5, which is fairly high considering the variability of teachers' estimates" (**16**, p. 71). Primary school teachers were found to know so little about the children in their classes as to make their ratings useless.

In this introductory summary of psychological tests in India, it would be remiss to omit the useful work of E. W. Menzel (**19**). Menzel is associated with an American mission school and is noted in India for his consultant work in connection with tests for secondary school use in the State of Madras. His useful book, summarizing much research with aptitude and achievement tests, has gone through three editions. He has done an excellent job of collecting scattered materials in a country where communication among research workers is very poor, but one reads this volume with disappointment because of the lack of validity data cited (because they are not available) and there is not a single mention of any attempt at cross validation. Menzel, however, has done an enormous service to India in presenting and summarizing this work, thereby making himself practically a one-man communication team for other interested workers in the field. He is also astute about the problem of translating foreign tests into local Indian languages, stressing the need for local norms, for detailed instructions to examinees (more detailed than needed in the United States), and for less emphasis on speed (20).

The lack of communication among research workers in India, along with the minute financial help that is available, is a little appalling to the outside observer. The annual Indian Science Congress schedules a few papers on psychology, but there is no national psychological association despite the many practitioners. Psychological journals are few and are apt to appear irregularly—again, finances—and do not receive the wide circulation that comparable journals do in the United States or Great Britain. The Southeast Asia Office of Unesco (in New Delhi) has begun in 1953 the much needed job of an annotated bibliography of all psychological articles published in Indian journals.

Of the journals, the oldest and best known is the *Indian Journal of Psychology* published by the Department of Psychology of Calcutta University. Formerly this was issued several times a year, but now it appears rather irregularly. *Shiksha* (published in Lucknow) and the *Indian Journal of Educational Research* (Bombay) also present material in the testing field. Other journals are largely devoted to education but, from time to time, do publish articles devoted to strict psychological concern: the *Journal of Education and Psychology* (published in Baroda) and the *Punjab Education Journal* (Jullundur) are two examples. *Teaching*, an excellently printed quarterly put out by the Oxford University Press, Bombay, while largely devoted to education problems, also carries significant articles in the field of psychology. An important newcomer on the horizon is the *Journal of Vocational and Educational Guidance* (Bombay), the first issue of which appeared in January, 1954. American readers will find most of these journals listed in *Psychological Abstracts*. All of them are in English, still the language of the universities and of academic communication generally.

Aside from these general remarks, it is easier to summarize the main developments in psychological testing and other areas by reference to specific locales which the writer visited during his year in India.

CALCUTTA

Department of Psychology, Calcutta University. Here the first psychological laboratory in India was organized in 1916 for graduate students under the direction of Dr. N. N. Sengupta after his return from Harvard where he studied under Munsterberg. The impetus toward the organization of the Applied Psychology Section was aided by the presence of foreign psychologists (Myers, Spearman, Jung) during the Silver Jubilee Session of the Indian Science Congress in 1938. For many years Dr. G. Bose was the head of the Department of Psychology and he brought an interest in clinical as well as other applied problems. With the retirement of Bose, Dr. S. C. Mitra (Ph.D., Leipzig) took over the chairmanship of the Department and through his efforts the Applied Psychology Section has been further strengthened. Students of the department may specialize in clinical, industrial-vocational, or educational psychology.

S. K. Bose (trained under Likert at

Michigan) is in charge of the Applied Psychology Section at the present time. The work of this section now consists of research with psychological tests, educational and vocational guidance, clinical examination of mentally deficient or emotionally maladjusted children, work in the field of industrial human relations, student training (**31**).

Vocational and educational counseling is available for both students and members of the outside community. A standard test battery is used, which takes about eight hours, and a standard fee is charged; for the clinical examinations, a higher fee is charged.

The Terman-Merrill Scales (L and M) have been adapted and are administered in Bengali; present versions are the result of item analyses with both forms and these 1952 revisions were being tried out on new samples (**31**, p. 5). Concrete intelligence is measured by a battery of performance tests (Alexander's Pass-along,[4] Dearborn Formboard, Kohs Blocks, and a special cube construction test developed locally). There is also a variety of special aptitude tests which will be familiar to American readers: the Stenquist Mechanical Assembly, adapted to local conditions; a nut and bolt test; a manual dexterity test; a test of constructional ability.

Tests of temperament comprise the following: (*a*) an inventory of 30 paired questions, the scoring of which provides a self-estimate of the examinee and which is based upon the Subjective Paired Word Test of the National Institute of Industrial Psychology in England; (*b*) a paper-and-pencil test of introversion-extroversion and a word association test, both locally developed; (*c*) a neurotic inventory prepared at Calcutta with clinical weights subjectively determined (**9**). Bhattacharya (**8**), one of the staff at Calcutta, has, however, published data to show that all of these indices have low reliabilities and are in need of revision. Both the Rorschach and TAT are being tried. The first TAT version was administered to some 200 students along with a standard personality inventory with a positive correlation resulting. Consequently a second set of Indianized pictures has been prepared by S. K. Bose and is now being tried.

For testing scholastic ability, three tests of reading, dictation writing, and arithmetic are employed. In addition, a physical and psychological examination (reaction time, sustained attention, steadiness, fatiguability) is conducted.

The Section in 1952 was engaged in the adaptation of the Wechsler-Bellevue scales (I and II) and in the determination of norms for Indian children for Raven's Colored Matrix Test. Age norms for different performance tests have been prepared and published (**31**) but only mean scores are supplied.[5] An engineering aptitude test and a clerical aptitude test have been constructed and are being standardized.

Individual members of the Department conduct special research projects. S. K. Bose has engaged in studies of employee attitudes and job satisfaction in several industries using the Likert-type scales (**10**). One of these employee attitude scales (it happened to be the Thurstone type) shown to the writer consisted of some 20 statements, printed both in Hindi and Bengali to ease the language problem; with illiterates, an interview is held with the worker in which all the topics on the scale are covered, the interviewer then filling out the scale at the end. The Ministry of Education, Government of India, has offered financial assistance for this work. Bose has also conducted a study of accident-prone tram and bus drivers under a grant from the Indian Council of Medical Research (**11**). S. P. Ghosh, a recent Calcutta Ph.D., has developed test batteries for semiskilled (manual) and skilled (mechanical) workers, criteria for which were ratings by supervisors as well as output. S. N. Roy (another recent Calcutta Ph.D.) has worked in the field of job analysis and was contacting successful people in medicine, law, engineering, accountancy, journalism, and teaching for information about training requirements and temperamental qualities needed for these jobs. S. C. Bisi was engaged in a job analysis of overseers connected with the engineering profession.

---

[4] An ingenious group version of the Pass-along has been developed by G. Bose.

[5] I saw very few cumulative percentile curves in India; "norms" typically refer only to average scores

Possibly because of its favorable location in Calcutta, the Section also conducts research on employee selection for various industrial firms. A general assessment battery is administered and candidates are then rated "A" to "D," such ratings being returned to the firm with recommendations (hire or not) from the Section. Follow-up studies have also been attempted.

Department of Anthropology, Government of India. Under the direction of Dr. B. S. Guha, this government agency is composed of several branches: physical and social anthropology, psychology, human biology, nutrition, statistics, X-ray. The Department is housed in spacious quarters in the Indian Museum in downtown Calcutta; there is an excellent library which contains many American psychological journals. The organization and history of the Department have already been described for American readers (**3**). Comments here will be restricted to psychological work. Psychological trainees are selected by various state governments. A certificate is given at the conclusion but no degree.

Dr. Uma Chowdhury (a recent Ph.D. from Calcutta) is in charge of most of the psychological work. Her main work has been in connection with a Department team, Unesco sponsored, which conducted a community study of an urban and village area in Bengal (Northeastern India). Chowdhury was responsible for the administration of the psychological tests used (TAT, Rorschach, Horn-Hellersberg, Wechsler-Bellevue, an ethical judgment scale). She reported that nearly all the Wechsler items needed to be changed and several revisions have been made. In late 1952, I checked through the recent changes made in this Bengali version. All the physical aspects of the revision (box, line drawings, photographs used) were well done. The test was not a mere translation of Wechsler items into Bengali. Vocabulary and Picture Arrangement were altered the most. Some 300 protocols for ages 8–25 were collected; an item analysis was contemplated as the next step before arriving at the final adaptation of the scale. There is still a shortage of adequate records for boys. The interesting comment was made that girls, after marriage, appeared to show "deterioration," i.e., they then cease to move in as stimulating an environment as before, when they were in school.

BOMBAY

Parsi Panchayet Vocational Guidance Bureau. This bureau is part of a larger social service agency known as the Offices of the Trustees of the Parsi

Panchayet Funds and Properties, an organization which conducts various welfare activities for the Parsi community (such as housing, medical aid, scholarship and financial aid, employment, nursery schools, home for the aged). The Vocational Guidance Bureau, established in 1947, is only one of these welfare activities; guidance services, however, are available to members of all communities rather than only Parsis. The Bureau is under the direction of Dr. H. P. Mehta (Ph.D., Columbia). It receives cases from various sources: from social workers attached to the parent body and from other Bombay social agencies, from psychiatrists and doctors for assessment of general intelligence of cases suspected of mental deficiency, from the Employment Bureau for a study of specific aptitudes of clients needing placement, from the nursery school in regard to behavior problem cases. Students in search of vocational guidance are generally self-referrals. The Bureau also conducts testing and guidance programs for boys and girls in the final class of certain Bombay high schools. Not surprisingly, it was reported that vocational guidance consultations were usually complicated by other problems, such as personal, social, economic difficulties. Fees are charged and are adjusted to ability to pay.

The Bureau lays emphasis on counseling and individualized work. Considerable time is devoted to each client. The general approach is nondirective. Tests are not administered in the form of a standard battery given to all clients; instead aptitude tests are used according to the needs of the situation. The vocational plan is jointly arrived at by client and counselor.

Most of the aptitude tests utilized are originally American, but they have been restandardized on local samples. In general, American norms have proven too "high" or else it has been necessary to extend the time limits.

A significant step forward has recently been taken by the Bureau in sponsoring the new *Journal of Vocational and Educational Guidance*. One of the great needs in India is that of communication among various workers in the field, and it is hoped that this journal will fill the present gap. The first issue contains two articles by the present writer on the vocational guidance situation in India (**4**, **5**), an article on the work of the Vocational Guidance Bureau of Bombay State (see below), an article providing national unemployment statistics from the Ministry of Labour, a reprint of Super's article (**30**) dealing with the self-concept of vocational adjustment, an occupational information piece on the legal profession, a general plea for more vocational guidance in the secondary schools taken from the Report of the Secondary Education Commission for the Government of India in 1953, an article dealing with the need for age-grade norms—all in all, a rather auspicious beginning. The editor is Dr. H. P. Mehta.

Vocational Guidance Bureau, Bombay State. Whereas the Bureau operated by the Parsi Panchayet Trust is a private agency, the State of Bombay established its own vocational guidance bureau in 1950 with the following objectives: (*a*) the collection and dissemination of occupational information, (*b*) the preparation of a cumulative record card for use in secondary schools which, as of the spring of 1953, has been sent to the Government of Bombay for approval; (*c*) the standardization of certain psychological tests for school use. The Bureau is under the direction of F. S. Chothia (Harvard trained).

Concentration has been largely directed toward the first objective, since this was felt to be the greatest need at the outset. Accordingly, the Bureau has published some seven pamphlets and prepared a series of 33 career monographs. Typical topics include social work, radio, engineering specialties, fisheries, police force, mining-metallurgy-geology. Career conferences have been arranged for both secondary school students as well as college students; these have been held in various cities throughout Bombay State. The Bureau has prepared a number of posters for school use; radio talks and lectures are given by the Bureau's staff. With the help of the Bureau, a documentary film on vocational guidance was produced by the Government of India Films Division and is now available for use by all interested organizations.

The Bureau realizes that effective work in vocational guidance can only be done if schools and colleges have a staff member trained in this specialty. With this in mind, the Bureau conducts what are termed Career Master courses. Since 1951 (and up to the spring of 1953) six conferences were held for secondary school teachers; in addition, one has been held for college personnel. Over 200 individuals have completed the courses, which consist of about 30 lectures in addition to some experience in psychological testing and visits to Bombay institutions. It is expected that when they return to their own schools they will set up Career Information Centers. To keep the alumni in touch with recent developments, a monthly newsletter is sent out.

In connection with the third objective, various psychological tests, mostly of American origin, have been adapted and restandardized. The Bureau is interested in doing follow-up studies with certain of these tests after they have been administered to entering first-year professional school students. A battery consisting of the DAT Abstract Reasoning, Bennett Mechanical, and Revised Paper Form Board has been given to students in engineering, architecture, and medical schools. Students at two technical high schools have also been tested. The data are too detailed for presentation here. Suffice it to say that the pretests have proved to be efficient predictors and, in some cases, better predictors than the usual entrance (essay-type) examinations. The plan is to follow up these students until the completion of their training.

PATNA

Department of Applied Psychology, Institute of Psychological Research and Service. This division of the University was until the fall of 1953 under the direction of Professor H. P. Maiti, who has recently been in the United States. A personal communication from him has informed the writer of a change in policy in relation to the Institute, with the result that many of its activities have been suspended and Professor Maiti has shifted to a new position at the Institute of Development at Ahmedabad. S. K. Mitra, who during 1952–53 was at the University of Chicago for training in clinical psychology, is apparently second in command. He appears to have been responsible for much of the research on the Rorschach (**21**) and other clinical tests at Patna. The following description of the work at Patna may have radically changed. What follows is a picture of activities as of the academic year of 1952–53.

The Institute was housed in its own building and was engaged in teaching, research, and service. In its teaching function, students received training for an M.A. degree in applied psychology (industrial or educational). The research and service functions were mostly confined to guidance and counseling, although there was some work in progress in therapy. The largest number of clients, both students and non-university people, comes for personal counseling and for vocational guidance. Student clients are largely at the B.A. stage and want to know what professional career is most suitable for them. The orientation of the Institute's staff was largely psychoanalytic and dynamic; consequently a good deal of stress was laid on clinical tests (especially projective ones). The common types of aptitude tests were also in evidence, but the sample on which norms were based was rather select: ages 16–21 years and generally composed of college students of the area who came voluntarily for testing and counseling.

A great deal of current research was under way in 1952 which can only be briefly indicated here. A test battery for Home Guard selection purposes was developed; it consisted of a verbal intelligence test in English, a psychomotor test of speed, digit-symbol test, TAT (local adaptation), and a group discussion situation. Some 1,700 candidates were tested from four districts of the province of Bihar. All summary test data were given to the Home Guard Board for the final selection of candidates. Another research project under way was a study of the 25 best and 25 poorest engineering students who were to be contrasted by means of a check list of the necessary traits for success, and, in addition, by a test battery including the Bennett Mechanical Comprehension, a test of spatial relations (similar to that of the National Institute of Industrial Psychology in England), a verbal intelligence test, and an achievement test. Another graduate student was working on a study of personality factors, using 8 TAT cards, with successes and failures in Grades VIII-X as judged by previous examination results. Still another student was engaged in a study comparing study habits of rural and urban school boys. Research was also under way with Rorschach (Indian norms for ages 10–16) and H-T-P tests as well as with a locally devised word association test. Continued research in connection with the Unesco Social Tensions Project was also under way. There were also some completed studies based on Sherif's work in perception and group norms (**29**).

LUCKNOW

Department of Philosophy and Psychology, Lucknow University. Professor Kali Prasad is the chairman of this department and, under his direction, the major emphasis has been on social and clinical psychology. There have, however, been studies in the field of industrial psychology and psychometrics.

The work of social psychology has been summarized in Murphy's recent book (**22**) and will not be repeated here; the department has continued with further studies in this area as well. Work along clinical lines has been directed toward psychodiagnostics and has been especially concerned with projective techniques (Rorschach, Picture-Frustration, Sentence Completion, a local version of the TAT). These have been used in connection with Unesco and Government of India studies on group tensions and social distance. Much of this earlier work has received a new impetus from the establishment of a psychological clinic dealing mostly with late adolescent and adult cases (Dr. Delton Beier of the Department of Psychology of Indiana University was Visiting Fulbright Professor during the 1952–53 academic year).

The major research in the field of industrial psychology has been a project concerning fatigue and efficiency in textile mills (26). The style and tenor of these reports is highly reminiscent of the earlier work of the (British) Industrial Fatigue Research Board. The Lucknow studies in this area have recently been extended to include research in accident causation.

Psychological testing, from the point of view of aptitude measures, has not been neglected. Such work, however, has not received major emphasis at Lucknow. Projects known to the present writer include an administration of Jalota's group mental ability test (see section on Banaras) and a master's project involving an original test to measure mathematical aptitude at the secondary school level. A Ph.D. student is working on an original group intelligence test in Hindi; another M.A. candidate was at work on applying the Porteus mazes to boys and girls of ages 5–15. Dr. H. S. Asthana (Chicago trained) has prepared a modification of the Thurstone Adjustment Inventory and has worked on Indian normative data for the group Rorschach.

Of interest to American readers will be the large-scale adaptation of the Cornell Value Study to large samples of Indian university students, data for which were planned by two visiting American social scientists in early 1953. Considerable and important comparative data about attitudes, opinions, family structure, occupational values, and the like will be available. Whereas the Cornell project involved only college males, the Indian questionnaire (made as nearly comparable as possible) will sample several thousand male and female university students.

Of all the psychological libraries that I inspected in India, Calcutta excepted, the one at Lucknow seemed to be the best. It contained an excellent collection of American psychological journals, twenty-five titles at least, most of which went back to volume one.

DELHI

Central Institute of Education. Largely because of its emphasis on teacher training (the general equivalent of the M. Ed. and Ed.D. degrees are awarded here), the Central Institute of Education has been concerned with aptitude and achievement tests for use in secondary schools. Projects by students have necessarily represented small-scale studies; only certain of these are abstracted in the Institute's publication, *Short Reports of Studies in Education and Psychology*. Certain faculty members have, however, been responsible for some of the more coordinated work in this area.

A Psychological Research Wing is also part of the Institute; a child guidance center has been created there with the help of Dr. James Caughley from New Zealand, who was brought to India by Unesco. During the 1952–53 academic year, the work of this center was extended to one of the Delhi hospitals and to a women's colony where refugee children were housed (and here also was started a class for retarded children).

Under the direction of Uday Shankar, the chief psychologist, a large research project involving the standardization of intelligence tests in Hindi is under way. Two research fellows and four testers comprise the additional staff. An individual scale of intelligence for ages 3–15, modelled on the Terman revision, is to be standardized. The test will consist of 128 items, divided into two forms, selected from a variety of sources and adapted to Indian conditions. It is proposed that 1,400 children, chosen randomly from 24 Delhi State schools, will be tested. Plans were that the entire testing of this sample would be completed by May of 1954, after which time validity studies were expected to get under way. A second project here is the standardization of a group test of intelligence for children between 5–18 years. Some 150 items for try-out have been collected and the actual testing was to commence in August, 1953. Some research has also been done in the area of selection of student teachers (**24**). The present writer revised and adapted the Minnesota Teacher Aptitude Inventory for use at this institution.

Psychological Research Wing, Defence Science Organization, Ministry of Defence, Government of India. This organization is largely concerned with the selection of Indian Air Force officers, all of whom are secondary school graduates at least; the work is based primarily on British experience during World War II although the staff is by no means unfamiliar with American research in this area. Security regulations prevent discussing many specific details of this program. The program of aptitude testing for each candidate consumes about three entire days. An enormous amount of psychological data is available that, it is hoped, will someday become available for civilian use in India. Of all the testing programs inspected by the writer, this was by all odds the best organized and the most adequately administered. The entire staff consists of about 20 persons who work in various sections: intelligence, aptitude, personality, training, statistics. Dr. Sohan Lal is the chief psychologist.

BANARAS

Department of Psychology and Philosophy, Banaras Hindu University. Under the present direction of Professor S. Jalota work is being carried on concerning group verbal tests of intelligence, both in Hindi and the English language. Dr. Jalota's emphasis on large student norm groups as well as age norms deserves attention. He has been much concerned with the frequent use of small and unrepresentative samples as well as the lack of validity data.

Work on the group mental ability scale was begun in 1951. The present version of this test in Hindi consists of 100 items (vocabulary, similarities, opposites, number series, reasoning, etc.) with a time limit of 20 minutes. It is the result of an item analysis involving a group of 200 students. Items are arranged according to difficulty values ascertained on this sample. Norms are available for educational levels equivalent to American high school and first-year college and for ages 12–17. The Manual of Directions (in Hindi) also provides a table for the conversion of raw scores into MA equivalents. Separate answer sheets are available, a rarity in India despite their economy. In 1952–53, the test was administered to a total of 2,200 additional students all in Hindi-speaking areas with little change in the original norm group data. Intercorrelations of subtests range from .3 to .6; correlations of subtests with total scores, from .7 to .8. A recent personal communication to the writer indicates that the test is not overloaded with items merely testing verbal facility in Hindi: correlations between test scores and class marks in Advanced Hindi are approximately .27 as is also the case with class marks in advanced mathematics (these results obtained on Allahabad student groups).

Similar work is being done on a

group test in English and Malayalam. Both of these versions have been item analyzed but final validity statistics have not been worked out.

MYSORE

Department of Psychology, Mysore University. Under the direction of Dr. B. Kuppuswamy, formerly at the University of Madras, this department has carried on extensive research on psychological tests in connection with its program of vocational and educational guidance. The research has involved certain general ability tests, aptitude and achievement tests, as well as personality and interest inventories. All are in English.

The performance scale of Collins and Drever is used as a general ability measure. Five hundred children have been tested and norms are available for ages 4–8. Reported reliabilities are not very satisfactory; validity, using the criterion of nursery school teachers' records, is .45. Alexander's Performance Scale for secondary school students has been administered to 300 students. Test-retest reliability for the entire battery is .86; validity, against school marks, is low (.23 for the entire battery). However, Cube Construction correlates .4 against this criterion.

In 1953 work was in progress on a verbal test of intelligence of the omnibus type for the high school level as well as a nonverbal intelligence test for student use. Plans were to standardize both of these during the 1953–54 academic year.

Certain manual dexterity tests have been used for appraisals of students in technical and high school curricula as well as with factory workers. Four of these are the familiar laboratory items of tapping, steadiness, precision, dynamometer; reliabilities, not surprisingly, vary widely; validity here has been checked against teachers' estimates of success in technical institutes for about 120 students, the coefficients varying from 0 to .6 (in the case of the precision test). Six more complex manual dexterity tests involving coordinated movements of both arms are also used. Examples are nut and bolt assembly, Minnesota Manual Dexterity, two-hand coordination. Some of these tests have been used on small groups of laborers; validity, as against supervisors' estimates, varies from .52 to .75 (for an *N* of approximately 50).

Two personality inventories have been developed. One of these is a test of temperament (cyclothyme-schizothyme) consisting of 80 items which has been administered to 650 college students. Split-half reliability is reported as .69; validity, using known groups, varies from .78 upwards for the individual traits. The second measure is a neurotic inventory, an adaptation of the Woodworth inventory, which has been administered to 200 college students and 200 secondary school pupils. Norms are available in terms of a five-point scale (well adjusted to poorly adjusted). Split-half reliability is .84; known normal and neurotic groups are sharply delineated.

At Mysore was possibly the best equipped psychological laboratory which I was privileged to see in India; furthermore, it seemed to be extensively used. (Kuppuswamy's Ph.D. thesis, incidentally, was a study in genetics using rats—the only mention of the use of animal data I encountered in India.) Two members of the staff have produced a good experimental psychology manual which contains instructions for some 50 simple experiments for undergraduate use (17). Topics include experiments relating to attention, imagery, memory, learning, reasoning, reaction time, conditioning, fatigue, etc. The appendix provides examples and formulas for simple statistical measures: mean, median, mode, *Q*, *SD*, rho. In the main the apparatus to be used are kept simple, utilizing as much paper-and-pencil material as possible.

Studies in the area of industrial psychology have not been neglected. An example is the work of Amin (2), who administered psychological tests to a group of 15 powerloom weavers (tests of threading, warp placing, pin placing, etc., developed on the basis of job analysis). Supervisory ratings collected some four months later, in addition to production data, correlated about .55 with the aptitude measures.

MADRAS

St. Christopher's Training College. In 1949 at the request of the Government of Madras, E. W. Menzel acted as consultant and advisor on aptitude tests for use in Madras schools. St. Christopher's Training College was selected as the center for this research. The result of this work is that four of Menzel's tests are now available for distribution free of charge to interested users: two intelligence tests and two achievement test batteries in the Tamil language for arithmetic, history, geography, and Tamil. There is also an English language test. All of these tests are designed for use approximately at the beginning of the high school course.

Under the direction of Dr. Jean Forrester research is being conducted on the selection of students for teacher training. Interviews, various situational tests, group discussions, and the Madras Senior Level Intelligence Test are used along with a medical examination. As of the spring of 1953, results from these had not been treated statistically.

Department of Psychology, University of Madras. Dr. G. D. Boaz, chairman of this department, may be familiar to some American readers since he was in America during 1949 and visited several university psychological clinics. Aside from the teaching aspects, interest in research is largely in clinical psychology. Boaz has recently submitted to the Ministry of Education an ambitious plan for the establishment of a mental hygiene clinic but the prospects for implementation appeared to be dim. Boaz does operate a private clinic on his own which is primarily for neurotics. To teach the advanced subjects in psychology (as experimental, mental measurements, social) the small staff at the University is increased by using faculty members from other colleges in the area. In early 1953 two research projects were under way. One of these, under the direction of Boaz, was the adaptation of Murray's TAT pictures to the university student population; another was a Ph.D. thesis on the problem of vocational education (a follow-up of graduates of technical schools) in Madras State.

ALLAHABAD

Bureau of Educational Research, Ewing Christian College. Dr. A. Edwin Harper, Jr., who has a Ph.D. from Columbia University, is in charge of this program. Ewing Christian College is one of the many Presbyterian mission schools scattered over India. (Dr. Harper during 1953–54 was a Visiting Associate of the Educational Testing Service, Princeton; upon his return to India in 1954 he plans to spend an additional year or two with Dr. P. C. Mahalanobis, the statistician, in Calcutta.) He brings to his work in India an excellent background both in statistics and psychology. He has therefore turned his research attention, admittedly limited because of mission school activities, to the development and standardization of certain psychological selection procedures for use in Indian schools.

Research in 1953 was under way with revisions for Indian use of the Iowa Physics, Chemistry, and Mathematics Aptitude Tests. Experimental use of these tests was to have commenced in April, 1953. Two tests are planned: a physics-chemistry aptitude and a mathematics aptitude, both in simple English. The tests are so designed that the material involving reading comprehension can be easily translated into the language of the region where used, if need be (**14**).

Dr. Harper has also prepared a Diagnostic Test of General Knowledge both in English and Hindi. Two forms are available, one for arts and one for science students. The test provides separate scores for general science, literature and language, social science, religion and sports, study skills, and background for understanding current events. The present version is the fourth revision of the tests. Reliability is above .90 for both forms.

A Hindi translation of the Bennett Mechanical and limited use of the Cattell Culture-Free Test have produced disappointing results.

A bilingual form of the Cornell Medical Index (questions printed in both English and Hindi) is also available. In a valiant attempt to control the use of this inventory, as well as to learn its usefulness "in the field," Dr. Harper has made the stipulation that copies will be made available only on the understanding that the completed forms will be returned for statistical analysis. Separate forms for men and women have been constructed.

Dr. Harper has also made small but laudable attempts to devise simple psychological experiments for undergraduates using local material. For example, to introduce students to tachistoscopic data, he has used English as against Hindi numerals and had students compare the results (English was clearly superior—and despite the unplanned lower brightness value of the printed cards).

Mention should also be made of Dr. Harper's consultant work at Vellore Christian Medical College (in South India). The college uses a two-stage selection process. First, aptitude and achievement tests are administered in centers all over India, Pakistan, and Ceylon. (The Moss Medical Aptitude Test was formerly used but now the college is developing its own measuring instrument.) On the basis of this preliminary screening, the 75 most promising candidates are brought to Vellore for the second stage. This consists of certain situational-type tests on the OSS model which were instituted by Dr. Frank Lake, formerly of the British Army. Group Rorschach has also been tried and good results obtained with an adaptation of Monroe's Inspection Technique. The selection board also considers the socioeconomic background of the candidates and the province from which they come; in certain cases this information will serve to counteract low aptitude test ratings. Dr. Harper has acted as a consultant since 1951 and has helped conduct research to evaluate the effectiveness of these new selection procedures as against the old "college marks plus recommendations" approach.

Bureau of Psychology. The Bureau is a government-sponsored agency established in 1947. It specializes in psychological testing, largely for school groups, and applications of clinical psychology. Dr. C. M. Bhatia is the director and his major research project has resulted in the Bhatia Performance Battery. Both individual and group guidance is available (7). Clinical tests in use are the Rorschach, "applied under local conditions" (32), an Indian version of the TAT, and a free association test in Hindi.

Five individual tests comprise the Bhatia battery: Kohs Blocks, Passalong, Picture Construction, Pattern Drawing, Immediate Memory. A factor analysis of these five tests shows two clear clusters: *g* plus a spatial factor and, secondly, a memory-verbal factor. The five tests are weighted to produce a performance quotient. Criteria employed were school records (for literates) and village opinion (for illiterates). Odd-even reliability is cited as .85 for the former and .84 for the latter. PQ validity was reported as .70 for 642 literates and .72 for 512 illiterates (6).

The Bureau has also developed and published several group verbal and nonverbal tests of general mental ability in Hindi. One of these is the Sohan Lal[6] Allahabad Intelligence Test. This is a spiral omnibus type test suitable for ages 11 and up, norms being available for age groups only. Split-half reliability is .94 for a group of about 1,400 individuals; validity had not been ascertained. Rugh (28), using this test with students 10–19 years of age, reports a correlation of .44 with mid-year grades. The distribution of scores he obtained was bimodal and, furthermore and probably more important, there was too much emphasis on the ability to follow written instructions. It was as if the instructions required a higher mental level than the items themselves.

The Bureau has also done extensive testing of candidates for the Police Training School, making the results available to the selection boards of such institutions.

From all appearances, the Bureau received very favorable treatment, especially financially, from the government of the State of Uttar Pradesh. Recently five branch centers have been established throughout the state, all under the direction of the Allahabad bureau whose staff numbered ten people. S. N. Mehrotra (Edinburgh trained), the statistician on the staff, has published a good deal in regard to test norms and standardization details, all of these articles appearing in *Shiksha*. From the clinical or psychotherapeutic point of view, the staff members appeared to be poorly trained; I was informed of one young man, with only an M.A. degree in philosophy, who was "practicing" psychoanalysis.

SUMMARY

Research and developments in the area of psychological testing in India as of 1952–53 have been summarized. The orientation in psychology is still largely British but there is currently a noticeable shift to the United States. There is a great deal of interest in aptitude and achievement tests for school use, in applied or industrial problems, in the clinical area, and in tensions research. Most departments of psychology are still very closely tied to philosophy; undergraduate course offerings in social science are sparse. Inadequate communication among research workers greatly hampers research progress, with the result that there is much overlapping of effort. There is the ever-present tendency to look for "quick solutions" and, on occasion, to adapt foreign psychological techniques too uncritically. Much solid groundwork has been done in the field of intelligence testing but with emphasis on the concept of the IQ. Little basic research in theoretical areas in psychology has as yet been attempted. Descriptive notes have been supplied concerning all the major psychological locales in India.

---

[6] Dr. Sohan Lal was formerly the director of the Bureau at Allahabad. He is now chief psychologist of the Psychological Research Wing, Ministry of Defence, Government of India.

REFERENCES

1. Alexander, W. P. *Intelligence, concrete and abstract.* London: Cambridge Univer. Press, 1935.
2. Amin, D. L. Experiment in selection of weavers on the basis of manual dexterity and further standardization of weaver's tests. *Indian J. Psychol.*, 1951, **26**, 105–109.
3. Anthropos. Department of Anthropology, Government of India. *Amer. Anthrop.*, 1948, **50**, 576–581.
4. Barnette, W. L., Jr. Vocational counseling in India. *J. voc. educ. Guidance, Bombay*, 1954, **1**, 9–12.
5. Barnette, W. L., Jr. Summary and recommendations of the Vocational Guidance Conference, Delhi, 1953, *J. voc. educ. Guidance, Bombay*, 1954, **1**, 39–43.
6. Bhatia, C. M. Mental survey of a village. *Indian J. Psychol.*, 1951, **26**, 111–115.
7. Bhatia, C. M. Educational and vocational guidance for students at the Bureau of Psychology, Allahabad. *Shiksha*, 1952, **5**, 62–72.
8. Bhattacharya, C. C. Comparative study of reliability of temperament tests. *Indian J. Psychol.*, 1950, **25**, 99–101.
9. Bisi, S. C. Study of neurotic questionnaire test. *Indian J. Psychol.*, 1950, **25**, 95–97.
10. Bose, S. K. Improving human relations in industries. *Industr. Relat.* (Indian Institute Personnel Management, Calcutta), 1952, **4**, 203–208.
11. Bose, S. K. Testing for accident-proneness in tram driver trainees. *Indian J. Med. Res.*, October, 1952.
12. Cox, J. W. *Mechanical aptitude.* London: Methuen, 1928.
13. Dayal, I. A note on psychological tests in India. *J. educ. Psychol., Baroda*, 1951, **8**, 195–197.
14. Harper, A. E., Jr. Adaptation of Iowa Aptitude Tests. *Teaching* (Oxford Press, Bombay), 1952, **25**, 45–46.
15. Jalota, S. *Text-book of psychology.* Banaras: Banaras Hindu Univer. Press, 1952.
16. Kamat, V. V. *Measuring the intelligence of Indian children.* London: Oxford Univer. Press, 1951.
17. Khadir, K. A., & Yoganarasimhiah, M. *Guide to experiments in psychology.* Bombay: Hind Kitabs, Ltd., 1952.
18. Mehrotra, S. N. Intelligence tests for selection purposes. *Shiksha*, 1952 **5**, 74–81.
19. Menzel, E. W. *Suggestions for the use of new-type tests in India.* (3rd Ed.) London: Oxford Univer. Press, 1952.
20. Menzel, E. W. Do's and don't's for intelligence tests in India and Pakistan. *Teaching* (Oxford Press, Bombay), 1952, **25**, 48–52.
21. Mitra, S. K. Evaluation of psychotherapy by means of the Rorschach test. *Patna Univer. J.*, 1952, **6**, 42–47.
22. Murphy, G. *In the minds of men: the study of human behavior and social tensions in India.* New York: Basic Books, 1953.
23. Naidu, P. S. *Research in education at the University of Allahabad.* Allahabad: Nandan Press, 1951.
24. Pires, E. A. Experiment in selection of student teachers. *Shiksha*, 1952, **5**, 22–32.
25. Prakash, R. Difficulties in adapting and translating mental tests in other cultures. *Indian J. educ. Res.*, 1952, **3**, 299–307.
26. Prasad, K. *Fatigue and efficiency in textile industry, Reports #1 and #2.* Lucknow Univer.: Laboratory of Experimental Psychology, 1950 and 1952.
27. Rice, C. H. *A Hindustani-Binet performance scale.* Princeton: Princeton Univer. Press and London: Humphrey Milford, 1929.
28. Rugh, D. E. Selection of high school students. *Christian Educ.* (Jahalpur, M. P., India), 1952, **29**, 20–25.
29. Sinha, D. Experimental study of a social factor in perception: the influence of an arbitrary group standard. *Patna Univer. J.*, 1952, **6**, 7–16.
30. Super, D. E. Vocational adjustment: implementing a self-concept. *Occupations*, 1951, **30**, 88–92.
31. Calcutta University, Department of Psychology, Applied Psychology Section. *Report on the working of the section*, 1943-1952. Calcutta: University College of Science, 1953.
32. *Bureau of Psychology, Uttar Pradesh.* India: U. P. Government, 1952.

Received April 27, 1954.

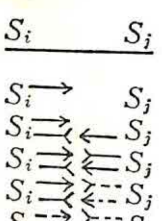
PSYCHOLOGICAL BULLETIN
Vol. 52, No. 2, 1955

ESTIMATING THE CHANCE EXPECTANCIES OF DIADIC RELATIONSHIPS WITHIN A GROUP[1]

RENATO TAGIURI, JEROME S. BRUNER, AND NATHAN KOGAN
Harvard University

The present paper discusses a mathematical model for computing the chance expectancy of diadic relationships defined by relational analysis procedures (**2**, **10**). In these techniques subjects are asked to select those group members they like most and least as well as to "guess" those who will like them most and least. The subjects may also be asked, in an extension of the method, to indicate whom they think each of the other members will choose and guess.

In relational analysis one of the basic units is the *diad*, i.e., the form of relationship that exists between a pair of group members in terms of whether they like each other and perceive each other as liking in return. The number of possible varieties of diads is either 10 or 45 depending upon whether one asks group members about their choices and perceptions of choice only, or whether they are asked to indicate their dislikes and perceptions of dislike as well.[2] A few of the 45 possible combinations of the latter case are shown below.



Legend:
→chooses
⇢rejects
—≺feels chosen
--≺feels rejected
a blank indicates omission or perception of omission

Interpretation of the observed frequency of the various diad forms requires that we know what the incidence of these diads would be if chance *alone* were operating, i.e., if group members were robots. We may then regard divergencies from these base-line frequencies as the result of *non*chance or systematic psychological factors. *The purpose of the present paper is to describe a model that provides an estimate of the chance frequencies and variances of each of the various diad forms.*

NATURE OF THE MODEL

Several chance models are conceivable. The one we choose corresponds to the case where the members are regarded as automata allocating their selections at random. The assumptions inherent in the model are these. First, independence is assumed between the different choices made by any single individual, and between his choices and his guesses. Second, the choices and guesses of any subject are taken to be independent from those of any other subject. Finally, there is the restriction that a subject may not choose or guess the same other person more than once.

The most general form of the model was developed by Luce and Macy and is applicable to the case in which the group members are permitted to

---

[1] This work was done in the Laboratory of Social Relations, Harvard University, under ONR Contract No. N5ori-07646.

[2] If the focus of attention is on the members rather than the diads, the number of different diads can be considered to be 81 or 16, instead of 45 or 10. For, when the analysis is concerned with the subjects themselves, it makes a difference whether S_i is related to S_j by a choice ($S_i \rightarrow S_j$) or by a choice-received ($S_i \leftarrow S_j$). From an impersonal point of view, on the other hand, the two diads are identical.

TABLE 1

Two Representations of the Diads Obtained When Choice Only Is Used

S_i		S_j			
C_i	G_i	C_j	G_j	S_i	S_j
0	0	0	0	S_i	S_j*
+	0	0	0	$S_i \rightarrow$	S_j
0	0	+	0	S_i	$\leftarrow S_j$
0	+	0	0	$S_i -\!\!\prec$	S_j
0	0	0	+	S_i	$\succ\!\!- S_j$
0	+	0	+	$S_i -\!\!\prec$	$\succ\!\!- S_j$*
+	0	+	0	$S_i \rightarrow$	$\leftarrow S_j$*
+	0	0	+	$S_i \rightarrow$	$\succ\!\!- S_j$
0	+	+	0	$S_i -\!\!\prec$	$\leftarrow S_j$
+	+	0	0	$S_i \substack{\rightarrow \\ -\prec}$	S_j
0	0	+	+	S_i	$\substack{\succ- \\ \leftarrow} S_j$
+	+	+	0	$S_i \substack{\rightarrow \\ -\prec}$	$\leftarrow S_j$
+	0	+	+	$S_i \rightarrow$	$\substack{\succ- \\ \leftarrow} S_j$
+	+	0	+	$S_i \substack{\rightarrow \\ -\prec}$	$\succ\!\!- S_j$
0	+	+	+	$S_i -\!\!\prec$	$\substack{\succ- \\ \leftarrow} S_j$
+	+	+	+	$S_i \substack{\rightarrow \\ -\prec}$	$\substack{\succ- \\ \leftarrow} S_j$*

* Symmetrical diads.

make as many choices and rejections, and to state as many guesses of choice and rejection as they wish (7). It can easily be simplified, however, to fit the case in which choices and guesses are fixed to a given and equal value by the investigator.

Let us take the case in which members of the group are asked to choose and guess a *fixed* number of others. Consider, now, the components of a diad formed by two subjects S_i and S_j: S_i's choice of S_j, or (C_i); S_i's guess of S_j, or (G_i); S_j's choice of S_i, or (C_j); and S_j's guess of S_i, or (G_j).

Each of these components may have a value of +, 0, or − indicating choice, omission, rejection, or their perception. There are 81 possible arrangements of +, 0, and − in the four components. Where rejections are not considered, there are only 16 arrangements. Since we shall be discussing the latter in detail, we present them in Table 1.

Notice that there are sixteen different entries but that some are only mirror images of others, thus leaving ten diad forms that are distinguishable, independently of whether they are "read" from S_i or S_j or from S_j to S_i.

Still another feature of the above array merits close attention. If we think of the choices (→) and guesses ($-\!\!\prec$) as bonds, it can be seen that diads can have 0, 1, 2, 3, or 4 such bonds joining its members. Values above four are not possible. The fact that diads can vary from 0 to 4 bonds is of central importance in what follows.

We may now state the general formula used to generate the chance expectancy of a diad form for the case in which choices and guesses are both fixed to a value d.

$$E=\frac{N(N-1)}{2}\;[P(C_i)P(G_i)P(C_j)P(G_j)] \tag{1}$$

Where:

N is the number of members in the group.

$P(C_i)$ is the probability of S_i choosing or omitting S_j, these probabilities being $d/N-1$ and $N-1-d/N-1$, respectively.

$P(G_i)$ is the probability of S_i guessing or not guessing S_j, these probabilities being $d/N-1$ and $N-1-d/N-1$, respectively.

$P(C_j)$ is the probability of S_j choosing or omitting S_i, these probabilities being $d/N-1$ and $N-1-d/N-1$, respectively.

$P(G_j)$ is the probability of S_j guessing or not guessing S_i, these probabilities being $d/N-1$ and $N-1-d/N-1$, respectively.

The variance for each of the estimates is obtained by the formula:

$$V_E = E\left[1 - \frac{2E}{N(N-1)}\right] \qquad (2)$$

As noted before, the number of bonds in a diad may be 0, 1, 2, 3, or 4. Now we shall show that in the case in which choices and guesses are both fixed to a value d there can be only five different values of E, the chance expectancy of the diad, and that these values are a function of the number of bonds in the diad. This is a result of the fact that the four probabilities $P(C_i)$, $P(G_i)$, $P(C_j)$, and $P(G_j)$ (cf. formula 1) can yield only five different products, since each of these probabilities can assume only two different values, $d/N-1$ for a choice or guess and $N-1-d/N-1$ for an omission, as indicated above. Now, if $P(C_i)$, $P(G_i)$, $P(C_j)$, and $P(G_j)$ can each have but two values, x for $d/n-1$ and y for $N-1-d/N-1$, there are only five different products that can be obtained by multiplying them: $(4x)$; $(3x)(y)$; $(2x)(2y)$; $(x)(3y)$; or $(4y)$. If x is the probability of a choice or guess, then the value $4x$ is the probability of a four-bond relationship involving choices and guesses by both S_i and S_j. If y is the probability of an omission, then $4y$ is the probability of a fourfold or mutual omission by S_i and S_j. The value $(2x)(2y)$ is that of a diad with two choices and/or guesses and two omissions, as, for example, where S_i and S_j both choose each other but omit guessing each other or where S_i chooses and guesses S_j but S_j neither chooses nor guesses in return. Thus the five probability products listed above, when multiplied by $N(N-1)/2$, (cf. formula 1), are the probabilities of 0, 1, 2, 3, and 4 bond diads.

There is one additional feature of the distribution of diads that must be clarified before we can proceed. Certain diads can occur in only one way, e.g., $S_i \rightarrow\leftarrow S_j$; others can come about in two ways, e.g., $S_i \longrightarrow S_j$ could also be $S_i \prec\leftarrow S_j$. That is to say, S_i can be a member of the latter kind of diad *by either choosing or guessing*, and so, too, S_j. In the first type of symmetrical diad $(S_i \rightarrow\leftarrow S_j)$ each subject can be a member of such a diad *only by choosing* (or in the case of $S_i \prec\succ S_j$, *only by guessing*). Thus, asymmetrical diads have twice the probability of occurrence of symmetrical ones.

COMPUTATIONS AND NOMOGRAPH

Case where data on rejection are not obtained. Let us turn now to the actual computation of chance expectancy for different diads where choice and its perception are to be studied and where data on rejection are not available. Using formula 1 for computing E, the chance expectancy, and formula 2 for computing its variance, we obtain the values given in Table 2. These are the expected frequencies and variances of any *symmetrical* diad with a specified number of bonds. When a diad is *asymmetrical*, however, its chance of occurrence is twice the obtained value since it may be formed in two ways, and, consequently, the values obtained from the appropriate formulas have to be doubled to yield the proper expected frequency and variance. The sum of the expected frequencies of all ten diad types must, of course, add to $N(N-1)/2$, the total number of diads in the group. Thus, for $N=10$, $d=3$, the frequencies are as presented in Table 3.

We showed above how the expected frequency of a certain diad with a given number of bonds is a function

TABLE 2

FORMULAS AND COMPUTATIONS FOR E AND V_E FOR A GROUP WITH $N=10$, $d=3$
CHOICE-ONLY CASE

No. of Bonds	Formula	Frequency of Each Diad	Variance
0	$\frac{N(N-1)}{2}\left[\frac{(N-1)-d}{N-1}\right]^4$	8.90	7.13
1	$\frac{N(N-1)}{2}\left[\frac{(N-1)-d}{N-1}\right]^3\left[\frac{d}{N-1}\right]$	4.44	4.00
2	$\frac{N(N-1)}{2}\left[\frac{(N-1)-d}{N-1}\right]^2\left[\frac{d}{N-1}\right]^2$	2.22	2.11
3	$\frac{N(N-1)}{2}\left[\frac{(N-1)-d}{N-1}\right]\left[\frac{d}{N-1}\right]^3$	1.11	1.08
4	$\frac{N(N-1)}{2}\left[\frac{d}{N-1}\right]^4$	0.56	0.53

of d and N (where choices and guesses are both fixed to a given value d). Indeed, under such conditions, it is the *ratio* of d to N that completely determines the chance of obtaining diads with 0, 1, 2, 3, and 4 positive bonds. The matter is made clear when one considers the extreme case where both choices and guesses are fixed at 4 and where group size (N) is five and the total number of diads, $N(N-1)/2$, equals 10. The significant ratio here is $d/N-1$. Where $d=N-1$, the ratio $d/N-1=1.0$. For this extreme case, four-bonded relationships are the only ones that can occur. Thus, the expectancy (E) of four-bond relationships is 100 per cent, and it is 0.0 for all other diad forms. At the other extreme, where N is very large compared to d, the ratio $d/N-1$ approaches 0.

This property of E as a function of $d/N-1$, when number of bonds is specified, permits the construction of a nomograph (Fig. 1) that can be applied to any $d/N-1$ ratio and thus obviate the necessity of using formula 1. The nomograph consists of five curves, corresponding to the five classes of diads with 0, 1, 2, 3, and 4 bonds. Note that the ordinate in the nomograph is a percentage of all possible diadic relations in the group. For example, if $d/N-1$ is equal to .5, the diads with two bonds

TABLE 3

THEORETICAL FREQUENCY OF THE VARIOUS DIAD FORMS

Diad Type		Value Obtained from Formula	Theoretical Frequency of Diad (E)
S_i	S_j		
( )	( )	8.90	8.90
→		4.44×2	8.88
—<		4.44×2	8.88
→ —<		2.22×2	4.44
→	←	2.22	2.22
→	>—	2.22×2	4.44
—<	>—	2.22	2.22
→ —<	←	1.11×2	2.23
→ —<	>—	1.11×2	2.23
→ —<	>— ←	0.56	0.56
			$\Sigma=45.00=N(N-1)/2$

12.870
9153

will constitute 40 per cent of all the diads in the group. To obtain the chance expectancy of any single diad, one decides first how many bonds the diad has. Then, one reads the percentage value from the curve appropriate to that number of bonds, entering the graph at the $d/N-1$ ratio. Finally, this value is divided by 1, 2, 3, or 6 depending upon the number of bonds and the symmetricality or asymmetricality of the diad (cf. inset table in Fig. 1). The variance can be computed by formula 2 after converting the percentage just found into a frequency by multiplying it by $N(N-1)/2$.

Case where data on both choice and rejection are obtained. Thus far we have been dealing with choices and guesses only. The case in which rejection as well as choice is used is only slightly more complicated, provided that the number of choices and rejections allowed, as well as that of their perceptions, is fixed at the same value d.

When choices and rejections are both used, there are 45 different diads possible, as noted earlier. As in the preceding case, expected frequencies are a function of N, of d, and of the number of bonds in a diad. Note, however, that here, too, there can only be diads of 0, 1, 2, 3, or 4 bonds. Consequently, even with rejections added, we find that the 45 possible diad arrangements can yield but five values for expected frequencies of diads of different numbers of bonds. A choice and rejection are mathematically identical when the number of choices (d_c) and rejections (d_r) have been fixed at the same value. The probability that S_i will choose or reject S_j must be $d/N-1$ as before. However, the probability of an omission is now decreased in view of the addition of the rejection category, becoming $N-1-2d/N-1$. If this value is substituted for the value $N-1-d/N-1$ wherever it appears in the computation formulas of Table 2, the theoretical frequencies for the choice-rejection case can be easily obtained.

The nomograph presented in Fig. 1 for the case where only choices are used is equally applicable to the choice-rejection case. Note, however, that one must use $2d/N-1$ (rather than $d/N-1$) on the abscissa for choice-rejection situations.[3] This doubling of the ratio is necessary here since members are allotting *twice* the number of selections as in the choice-only case.

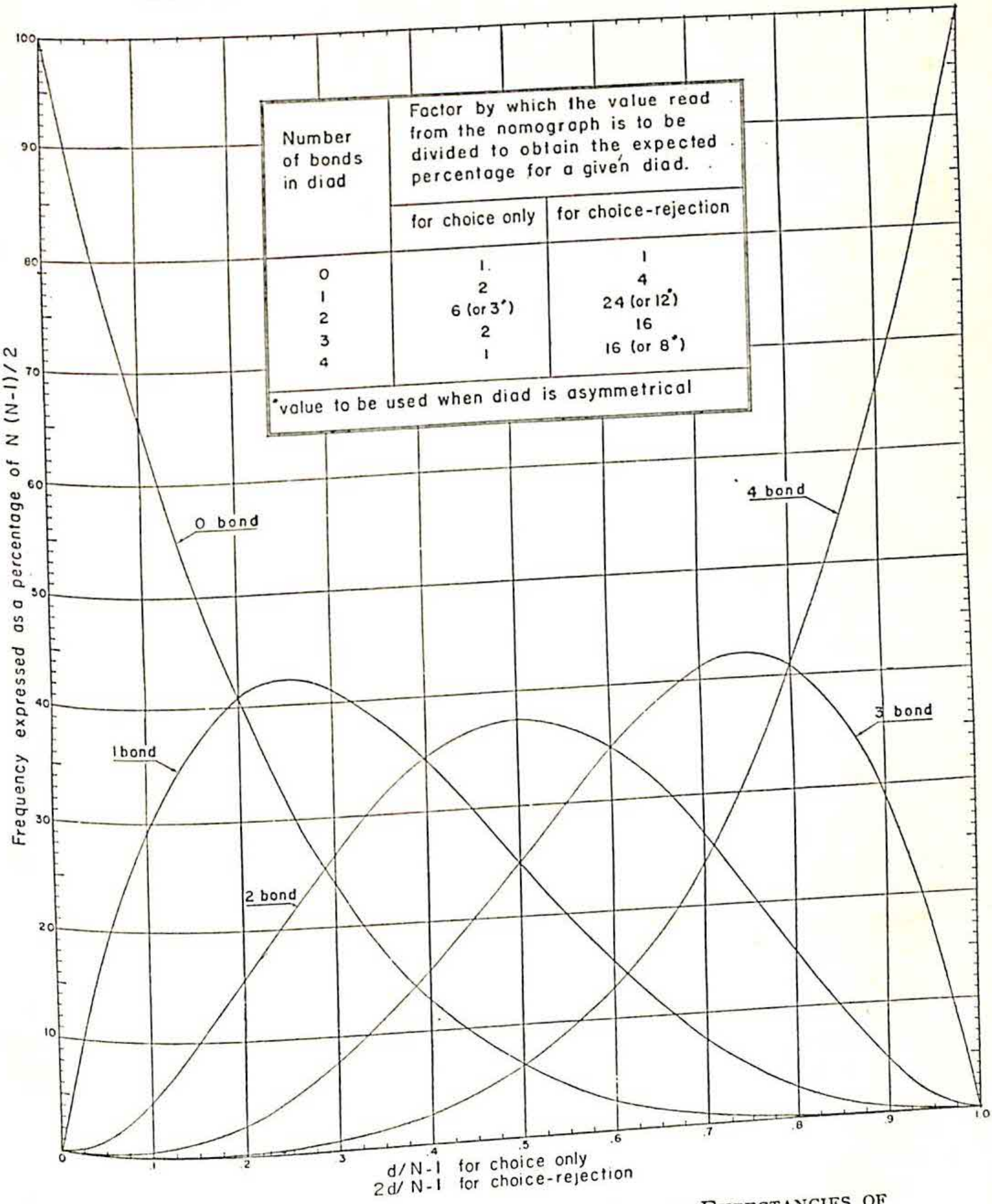
COMPARISON OF OBTAINED AND THEORETICAL VALUES

It is of interest to see how actual data compare with theoretical sets. In the example given in Table 4 we see that empirical data collected on two well-acquainted groups of Navy men ($d=3$ for each group) differ from chance for the 16-man group, but not for the group of six men. In the latter case both the observed and the chance frequency of 3- and 4-bond diads are relatively high. Thus, it becomes impossible to separate the influence of chance from the psychological factors at work. If, on the other hand, the investigator had used a $d=2$ in the six-man group, considerably smaller chance expectancies for multi-bonded diads would have resulted. The operation of psychological factors could then be more easily distinguished from the influence of chance. It is thus desirable to keep the $d/N-1$ ratio from becoming too large.

It is, of course, the discrepancy

---

[3] It must be remembered that when using both choices and rejections symmetricality includes the positive or negative nature of the selection or guess. Thus $S_i \rightleftarrows \leftleftarrows S_j$ is *not* symmetrical while $S_i \rightleftarrows \dashleftarrow S_j$ is.



Number of bonds in diad	Factor by which the value read from the nomograph is to be divided to obtain the expected percentage for a given diad.	
	for choice only	for choice-rejection
0	1.	1
1	2	4
2	6 (or 3*)	24 (or 12*)
3	2	16
4	1	16 (or 8*)

*value to be used when diad is asymmetrical

FIG. 1. NOMOGRAPH FOR ESTIMATING THE CHANCE EXPECTANCIES OF DIADIC RELATIONSHIPS WITHIN A GROUP

between expected and obtained frequencies that holds most interest for us psychologically. For example, in the 16-man group the tenth diad in the column greatly exceeds the expected value, while the sixth diad is below expectancy. The former diad in question is one that is marked by mutuality of choice and perception. The latter is one-sided, one member likes the other and this feeling is perceived but not reciprocated. The investigator might want to know whether the abundance of the former diad and dearth of the latter, relative to expected values, are statistically significant. He can test for significance since the model supplies him with the variance for the expected frequency of *each* diad (cf. formula 2). In this case, both types of diads differ significantly from the expected values. For the tenth diad, $E=0.1$; $V_E=0.1$; $obs.=8$; $t=25.0$; $p \ll .001$. For the sixth diad, $E=6.1$; $V_E=5.8$; $obs.=0$; $t=2.54$; $p<.05$.

We might want to know, moreover, whether mutuality of feelings in the 16-man group exceeds chance. By adding together those diads in Table 4, in which choice is mutual as well as those where omission is mutual, we find that their combined frequency exceeds chance significantly ($\chi^2=4.15$, $p<.05$, $df=1$). If mutuality of choice *only* is tested the discrepancy from chance increases ($\chi^2=6.27$, $p<.02$, $df=1$).

We can also study whether subjects are accurate in their perceptions of others' choices. By summing over all diads containing accurate perceptions, we find that accurate guesses exceed chance very significantly ($\chi^2=15.0$, $p<.01$, $df=1$). Similarly we might ask whether a member's feelings for another and his perception of the other's feelings correspond in excess of chance—which they do ($\chi^2=40.45$, $p<.01$, $df=1$).[4]

APPLICATION TO THE INDIVIDUAL CASE

So far, we have discussed the model as it applies to the expected frequencies of diads for the group as a whole. The investigator may be interested, however, in knowing what the expected frequencies and variances of each diad would be for an *individual member* of the group. Both values can be readily obtained. The expected frequency (E_i) of a specific diad for an individual S_i equals E/N, where E is the value expected for the group for that type of diad. The variance of E_i equals V_E/N, where V_E is the variance of the E for the group.

APPLICATION OF MODEL TO EXTENDED RELATIONAL ANALYSIS

The mathematical model discussed here can also be applied to certain problems encountered in an extended form of relational analysis, where group members indicate not only their own feelings and perceptions, but also have the task of predicting whom each of the other members of the group will choose and guess. Thus one obtains the distribution of the *actual* diadic relationships in the group as well as each *member's perception* of this distribution of diads. If, in this procedure, choices and guesses are both fixed to a value d, each of these distributions has the same probability characteristics. Thus, the two sets of distributions (actual and perceived) can be easily compared in a number of ways.

Consider an example of the use of the model for this case. Let us ask first whether the average member's conception of the relationships between the pairs of members in his group is different from what is expected by chance. Table 5 contains the relevant data, gathered from a 16-man group of Navy enlisted men. A chi-square test results in a value of 520.75 with a $p<.001$ ($df=9$). When the last diad ($S_i \rightleftarrows \leftleftarrows S_j$) is omitted in view of its extremely low expected value, chi square equals 30.75, still with a probability of less than .001 ($df=8$). We can then conclude that the two distributions differ and that the perceived relationships are not the result of haphazard guessing on the part of the members.

---

[4] For an extended discussion of these and related substantive problems cf. Tagiuri, Blake, and Bruner (11).

TABLE 4

A COMPARISON OF ESTIMATED AND OBSERVED FREQUENCIES OF THE VARIOUS DIAD FORMS (WHEN $N=6$ AND 16, $d=3$)

Diad Form	S_i	S_j	$N=6$			$N=16$		
			Est.	Obs.	χ^2	Est.	Obs.	χ^2
1.	( )	( )	0.4	2	6.40	49.1	74	12.63
2.	⟶		1.2	1	0.03	24.6	12	6.45
3.	—<		1.2	1	0.03	24.6	8	11.20
4.	⟶ —<		1.7	2	0.05	6.1	9	1.38
5.	⟶	⟵	0.8	0	0.80	3.1	1	1.42
6.	⟶	≻—	1.7	0	1.70	6.1	0	6.10
7.	—<	≻—	0.8	0	0.80	3.1	1	1.42
8.	⟶ —<	⟵	2.6	4	0.75	1.6	1	0.23
9.	⟶ —<	≻—	2.6	3	0.06	1.6	6	12.10
10.	⟶ —<	≻— ⟵	2.0	2	0.00	0.1	8	624.10
		Σ	15.0	15 $\chi^2=10.62$ $df=9$ $p<.40$	10.62	120.0	120 $\chi^2=677.11$* $df=9$ $p<.001$	677.11

---

* With the last diad form removed (because of its extremely low expected value) chi square equals 52.77, $p<.001$, df 8.

Now let us inquire about the degree of correspondence between the *actual* distribution of diads and the *perceived* distribution. Comparison of the actual and perceived (average) distributions (Table 5) yields a negligible chi square of 7.12, with a $p>.50$ ($df=8$). It is thus possible to say that the distribution of diads perceived by the members does not differ from the actual one but that it does differ from the chance base-line distribution. We may take this to be indicative of accuracy in the perception of group diads.

Though the group consensus corresponds to the actual distribution of diads, the question may be raised about the accuracy of some particular *individual*. Consider S_i whose data are presented in the last column of Table 5. His distribution differs significantly from the actual one ($\chi^2=33.39$, $p<.001$, $df=8$), but his performance cannot be attributed to random guessing in view of a significant difference between his individual conception and the estimated (random) distribution ($\chi^2=169.75$, $p<.001$, $df=9$). The latter chi square becomes 17.65 with $p<.05$ ($df=8$) even when the last diad form is not included because of its extremely low expected value. His error, therefore, is systematic rather than random. More frequently than not, error by individuals is found to be of this systematic kind, rather than being attributable to haphazard guessing. In S_i's case, for example, the data in-

TABLE 5

FOUR DISTRIBUTIONS OF DIADS: ESTIMATED (E), ACTUAL, PERCEIVED BY THE AVERAGE MEMBER, AND PERCEIVED BY A PARTICULAR MEMBER S_i

S_i	S_j	Estimated	Actual	Average Perceived	S_i Perceived
( )	( )	49.1	74	70.4	64
→		24.6	12	11.8	14
—<		24.6	8	11.2	15
→ —<		6.1	9	8.7	7
→	←	3.1	1	1.8	2
→	≻—	6.1	0*	1.7	5
—<	≻—	3.1	1	1.6	4
→ —<	←	1.6	1	2.7	4
→ —<	≻—	1.6	6	3.0	1
→ —<	≻— ←	0.1	8	7.1	4
		120.0	120	120.0	120

* When the actual distribution served as the "expected" in the chi-square analysis, this particular diad was omitted, yielding 8 degrees of freedom in all.

dicate that he differs from the rest of the members in underestimating the frequency with which fully reciprocated relationships ($S_i \substack{\rightarrow \\ \prec} \substack{\succ \\ \leftarrow} S_j$) exist in the group.

RELEVANCE TO TRADITIONAL SOCIOMETRY

Is there any tie between the model discussed here and those developed in traditional sociometric work? Several writers have proposed methods for obtaining the chance expectancy of *choices or rejections received by individual members* of a group by the use of Monte-Carlo methods (**8**) and the binomial expansion (**1**). The problem of how many *isolates* could be expected in a given group has also been dealt with in the same manner (**1, 4, 5, 8**). Investigators who have studied "*cleavage*" have concerned themselves with the expected volume of *choices exchanged between subgroups* (**1, 3, 4, 6, 9**), utilizing chance expectancies based upon empirical proportions of one subgroup to another.

The present model is tailored to a form of analysis that has the *diad* as its basis, and therefore does not have direct bearing on the above work. The diadic approach need not, however, be limited to relational analysis and can be profitably applied to any set of standard sociometric data. Such data yield three or six different diads depending upon whether choice only or choice and rejection is used. The chance frequencies of these six sociometric diads are readily obtained by applying formula 1, after removing the probability factors for the guesses. Previous investigators have concentrated primarily upon the estimate of chance frequency of one type of diad, the *mutual choice*. Proctor and Loomis (**9**), for example, show that the chance frequency of *mutual* diads with fixed d is given by $[N(N-1)/2]\ (d/N-1)^2$. A comparison of the above expression with formula 1 will indicate that they are of basically the same form, except for the probability factors involving guesses. This type of model represents the sociometric work most closely related to the type of diad analysis and model discussed here.

SUMMARY

This paper discusses the use of a mathematical model for computing the chance frequency and variance of the diadic relationships obtained with relational analysis. This procedure defines the diads within a group in terms of the pattern of choices, rejections, and perceptions of choice and rejection between pairs of members. The number of possible combinations is 45 when using choices, rejections, and the perceptions of these, and 10 when using only choice and perception of choice.

Evaluation and interpretation of observed diadic frequencies require knowledge of what the chance expectancy would be if the members of the group operated as robots. The mathematical model discussed here provides this necessary base line in a rational form rather than by recourse to empirical or Monte-Carlo methods.

Formulas for computing the expected frequency and variance of any diad type are given for the case in which choices, rejections, and their perceptions are all fixed by the experimenter at the same value, *d*. In this case, the expected frequency of any diad with a given number of bonds is a function of the ratio $d/N-1$. This property of the expected frequency of a diad permits the construction of a nomograph by which expected frequencies can be quickly calculated without resorting to formulas.

The model can be applied to the problem of whether an individual member is embedded in a diadic web that differs from what would be expected by chance. The model can also be used in connection with an extended form of relational analysis adapted to the study of perception of the diadic network in a group.

Examples are given of the application of this model to various problems requiring the comparison of observed data against a theoretical base line.

REFERENCES

1. Bronfenbrenner, U. The measurement of sociometric status, structure and development. *Sociometry Monogr.*, 1945, No. 6.
2. Bruner, J. S., & Tagiuri, R. Annual Technical Report No. 2, ONR Contract N5ori-07646. Laboratory of Social Relations, Harvard Univer., March 1, 1953, mimeographed.
3. Criswell, Joan H. Sociometric methods of measuring group preferences. *Sociometry*, 1943, 6, 398–408.
4. Edwards, Daisy S. The constant frame of reference problem in sociometry. *Sociometry*, 1948, 11, 372–379.
5. Katz, L. The distribution of the number of isolates in a group. Series 36, Institute of Statistics, Univer. of North Carolina, 1950, mimeographed.
6. Loomis, C. P., & Pepinsky, H. B. Sociometry, 1937–1947: theory and methods. *Sociometry*, 1948, 11, 262–286.
7. Luce, R. D., Macy, J., & Tagiuri, R. A statistical model for relational analysis. *Psychometrika*, in press.
8. Moreno, J. L., & Jennings, Helen H. Sociometric measurement of social configurations. *Sociometry*, 1938, 1, 342–374.
9. Proctor, C. H., & Loomis, C. P. Analysis of sociometric data. In Marie Jahoda, M. Deutsch, & S. W. Cook (Eds.), *Research methods in social relations: with special reference to prejudice.* Vol. 2. *Selected techniques.* New York: Dryden, 1951. Pp. 561–585.
10. Tagiuri, R. Relational analysis: an extension of sociometric method with emphasis upon social perception. *Sociometry*, 1952, 15, 91–104.
11. Tagiuri, R., Blake, R. R., & Bruner, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, 48, 585–592.

Received February 24, 1954.

Psychological Bulletin
Vol. 52, No. 2, 1955

THE ADAPTATION AND USE OF KENDALL'S COEFFICIENT OF CONCORDANCE (*W*) TO SOCIOMETRIC-TYPE RANKINGS[1]

BEN WILLERMAN
Student Counseling Bureau, Office of the Dean of Students, University of Minnesota

Suppose the members of a group are asked to rank all the members including themselves according to some criterion such as social influence or cooperativeness. It may be desirable to know how much agreement is present in the rankings. Rho's could be computed for the $n(n-1)/2$ pairs of rankings and an average rho would represent the extent of agreement. This becomes time consuming when the N of the group gets large.

Kendall's W (**2**, **3**) is an ideal measure for this situation and it is relatively rapid to compute. To obtain W, the ranks received by each individual are summed and the squared deviations of these sums from the mean computed. This quantity is then divided by the sum of the squared deviations which would result from perfect agreement among the rankers.

W requires that all members be ranked. But in sociometric situations this would mean that the rank a member assigns to himself would have to be included or that he be assigned some rank. If either of these procedures were followed, the index of agreement would be difficult to interpret. It would, therefore, be desirable if a W could be computed which would omit the ranker's self-rank.

The purpose of this paper is to present a formula for computing W when self-ranks are omitted. In the usual case W is defined as $12S/m^2(n^3-n)$ where S is the obtained $\Sigma(X-M)^2$, m is the number of rankers, and n is the number of objects ranked. We shall deal with the symmetrical case in which the number of rankers is equal to the number of individuals ranked. Our task is to find the quantity by which S is to be divided when self-ranks are omitted.

Let us first look at a model of a matrix of perfect concordance under this condition where the individuals are arranged in order of received rank.

	Rankees			
Rankers	A	B	C	D
A	—	1	2	3
B	1	—	2	3
C	1	2	—	3
D	1	2	3	—
Sum of ranks	3	5	7	9

It will be noticed that each individual's sum of ranks is different from his neighbors' by a constant amount. This difference will always be equal to $n-2$, because $n-2$ individuals will assign a given individual one rank larger than they ranked his predecessor, while the remaining two persons' ranks cancel one another, so to speak. These individuals are the rankee and his immediate superior. The superior assigns the rankee the same rank as the rankee gave him. It will also be observed that the first ranking individual's sum of ranks is always equal to $n-1$. These regularities enable a substitute formula to be achieved.

---

[1] This investigation was supported by a research grant (M-454C) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. It is also a product of a program of research in student life carried out by the Office of the Dean of Students, University of Minnesota.

To obtain the $\Sigma(X-M)^2$ we shall use the formula

$$\Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

with the following definitions: n = the number of individuals in the group, $m=n-1$, and $k=n-2$.

Rank	X	X^2
1	m	m^2
2	$m+k$	$m^2+2mk+k^2$
3	$m+2k$	$m^2+4mk+4k^2$
.	. .	. . .
.	. .	. . .
.	. .	. . .
.	. .	. . .
n	$m+mk$	$m^2+2mmk+m^2k^2$

$\Sigma X = nm + k\Sigma$ 1st m integers,

$\Sigma X^2 = nm^2 + 2mk\Sigma$ 1st m integers $+k^2\Sigma$ 1st m squares.

It can be shown that $\Sigma X = mn^2/2$ and that $\Sigma X^2 = (1/6)(2nm^4+3nm^3+nm)$.

It can also be shown that

$$\Sigma(X-M)^2 = \frac{nm}{12}(m^3-3m+2)$$

which is the substitute divisor for S.

The new W is therefore

$$= \frac{12S}{nm(m^3-3m+2)}$$

or

$$\frac{12S}{n(n^2-1)(n-2)^2}.$$

To test the hypothesis of independence among rankings the writer has prepared Table 1 showing the W's necessary to attain significance at the 5 per cent and 1 per cent levels for n = four through fifteen. The formulas used were taken from an article by Durbin (1) who has worked out a general solution to the incomplete block method in ranking experiments of which the present solution is a special case. Of the two methods he presents for approximating the distribution of W, the Beta distribution was used in preparing the table.

TABLE 1

CRITICAL VALUES OF W FOR THE 5% AND THE 1% LEVELS OF SIGNIFICANCE

n	5%	1%
4	.86	.95
5	.59	.73
6	.44	.55
7	.35	.44
8	.28	.36
9	.24	.30
10	.21	.26
11	.18	.22
12	.16	.20
13	.14	.18
14	.13	.16
15	.13	.15

REFERENCES

1. DURBIN, J. Incomplete blocks in ranking experiments. *Brit. J. Psychol.*, (Stat. Section), 1951, 4, 85–90.
2. KENDALL, M. G. *The advanced theory of statistics*. Vol. 1. London: Griffin, 1947.
3. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.

Received April 26, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 2, 1955

COMPLEX TASKS FOR USE IN HUMAN PROBLEM-SOLVING RESEARCH[1]

WILBERT S. RAY
Skill Components Research Laboratory, Air Force Personnel and Training Research Center

Progress in any field of research depends on many things in addition to the ideas produced by the individual research men, among them the availability of measuring instruments and of standard materials and techniques. The current high level of activity in research in learning theory, for example, would be impossible without the existence of nonsense syllables, mazes, conditioning techniques, and so on. The purpose of the present paper is to consider the extent of the use of such materials in the field of human problem solving, or, more specifically, in that part of the field which works with complex apparatus problems, by examination of a number of such problems which have been used in past research.

Development in the direction of standardized materials is already taking place in some sorts of human problem-solving and reasoning investigations: in work on concept formation, the Weigl and Vigotsky materials and the more recent Wisconsin Card Sorting Test and Hovland's flowers; water jar problems, anagrams, etc. in research on set; short, statistically evaluated problems of many kinds in factor analysis studies. In order to keep the present endeavor within reasonable limits, of the many remaining sorts of problems, only the more complex have been considered. Complexity refers here to the amount of work the subject must do in solving the problem: (*a*) several responses are usually required, (*b*) more than one hypothesis is available, (*c*) trial and error includes more than one or two trials, and (*d*) more time than that needed for a single brief act is required. Although complexity was the principal criterion, others were also used: (*a*) publication of a clear description of the problem, (*b*) the presence of a solution which could be classified as successful, and (*c*) in most cases, the presence of some sort of physical equipment in the task. Certain kinds of complex problems have been omitted on what may be a purely arbitrary basis, as when mathematical problems were ignored because their great number discouraged any attempt at selecting a few as samples. The same thing is true of card tricks and parlor puzzles of the bent nail variety. An additional criterion operated in the selection of problems in which the experimenter could watch the subject's solving processes at work, as opposed to problems where he sees only the solution, the end product.

Problem solving may be defined as the process of changing a given situation to a specified different situation, the process being new to the solver, and ending when the solution is achieved—which means that the process is not repeated. This definition will include cases in which the problem is solved by "pure" trial

---

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland AFB, San Antonio, Texas, in support of Project 7709. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

and error, as in McGeoch's Problem Box (task 28). If one wished to exclude that condition he could add: the situation being such that there is some basis for selecting one alternative as more probable than the others.

Following Duncker (**10**, p. 1), problem-solving tasks may be thought of as divided into three parts: (*a*) the given situation, (*b*) the desired situation, and (*c*) the method of proceeding from the one to the other. The individual problem-solving experiments discussed here seem to use either of two procedures; they ask the experimental organism to discover the operations which will change the first into the final situation, or they give him the first situation and the operations and ask him to predict the final situation. The above definition is intended to include both of these procedures. Most of the tasks described use the first of them. Székely's Candle Lever (task 1, below) and his Hydrostatic Levers (tasks 3 to 5) are examples of the second. Discovering the operations sometimes includes a fair amount of intellectual work; so one may classify here the form of the Candle Lever in which the subject is shown the first situation, the operations, and the final situation, and is asked to explain why the operations produced the result, i.e., is asked to provide the rational basis for the operations.[2] A different kind of example is provided by one group which worked on Maier's Hat Rack (task 10). The experimenter required this group to watch him build the hat rack and then asked them what possible function the structure could serve (**21**, pp. 352–353). Referring back to the definition of problem solving given above, "changing . . . to a specified different situation," it is the "specified" part which the hat rack subjects were asked to supply; so this example may be classified with the procedure of predicting the final situation.

The tasks considered in the preparation of this paper are listed and briefly described in the next section. They include all those which met the criteria described in a previous paragraph, and which were reported in articles covered in the *Psychological Abstracts* from the middle of 1948 to the middle of 1953, plus those in journals too new to have been abstracted by July 1, 1953. Actually, this search of *Psychological Abstracts* unearthed very few problem-solving tasks which had not been described previously. They and a few added older tasks have been referred to under the earliest references available for each. It is thought that the list includes most of the tasks of the desired sort which have been described in the psychological literature in the last twenty years.

---

[2] It is tempting to add further examples of providing an explanation from the demonstrations used by Piaget (**27**) and Deutsche (**9**), such as inverting a glass jar over a burning candle and asking why it stopped burning, putting a stone into a beaker of water and asking why the water level rose, or putting a penny into a box which has strings attached to the corners, swinging it in a circle overhead, and then asking why the penny does not fall out. These descriptions are not included, however, on the ground that they were not used as problem-solving tasks as problem solving is defined here.

As used by Deutsche (**9**) the problems present examples of several of the desirable attributes listed; they provide scoring continua since it was possible to classify the answers on an eight-point scale; and there is some evidence of validity since the split-half "reliability" coefficient of .74 suggests that if the experimenter has correctly identified some of the tasks as examples of some particular psychological process, then the others are probably examples of the same process. On the other hand, with regard to the "special knowledges" criterion, it seems not impossible that the answers the subjects gave to the questions might be better classified as knowledge rather than as reasoning, which does not imply that they are not examples of how children think.

DESCRIPTIONS OF THE TASKS

1. *The Candle Lever* (Székely, **32**)

Given and required: The central part of the task is an arrangement of a lever with a fulcrum toward one end of it, the lever being balanced by a candle on the short end. In one form of the task, the candle is burned, the lever tips up, and the subject is asked to explain why. In another form, the subject is shown the equipment with the candle not burning and asked what would happen if it were lit. In the third form, the parts of the apparatus are placed on the table with other objects, including a box of matches, and the subject is asked to construct an arrangement in which the lever will tip after a few minutes. In a variant of this third form of the experiment, the objects on the table include a labeled bottle of ether or benzine.

2. *The Inclined Plane* (Székely, **32**)

Given: A small cart stands halfway up an inclined plane, a string running from the cart over a pulley at the top of the plane and down to a metal weight. The weight is half submerged in a container of water so it balances the cart. On the table are a pipette and another container partly full of water. In another form of the task, the metal weight is replaced by a block of wood.

Required: Think of a method by which the cart can be moved up the incline 1 or 2 cm. Use the pipette, but do not touch the cart itself.

3. *Hydrostatic Lever, Single Weight* (Székely, **32**)

Given: With the subject watching, a container half full of water is placed on one end of a long narrow board (50×10 cm.) and the board is balanced on a fulcrum. A metal weight is hung just above the water. In a second form of the task, the metal weight is replaced by a wooden one of the same size.

Required: If the weight is lowered until partly submerged, not touching the sides or bottom of the container, what will happen to the lever?

4. *The Hydrostatic Lever, Symmetrical* (Székely, **32**)

Given: A lever with a metal weight at one end and a wooden weight at the other large enough to balance the metal when the lever is supported at the center. The wooden weight is higher and wider than the metal, and they are hung with bottom edges level. The weights hang in a large container which contains water up to a level just below the bottom of the weights.

Required: Suppose we put in more water so that both weights are immersed to a depth of 2 cm. What will happen?

5. *The Hydrostatic Lever, Unsymmetrical* (Székely, **32**)

Given: A lever with a metal weight at one end and a wooden weight of equal size at the other is balanced on a fulcrum, which would be closer to the metal weight. The weights hang in a large container which is filled with water almost up to the bottom edges of the weights.

Required: Suppose we filled the container with water. The rising water level would simultaneously touch the bottom of both objects. Now we let more water run in, until both objects are immersed by about 2 cm. What will happen? (**32**, p. 16).

6. *The Hollow and Solid Spheres* (Székely, **33**)

Given: After teaching the subject something of moments of inertia and of rotation, the experimenter says to the subject, "Imagine that we have two metallic spheres of exactly equal size and exactly equal weight before us. The spheres are covered with a lacquer which makes it impossible to distinguish the spheres from one another. However, there exists a concealed difference between the two spheres. One consists of a very light, the other of a very heavy metal. How is this possible?" (**33**, pp. 392–393). If the subject is not able to tell that one must be hollow, the experimenter tells him so. The experimenter then draws cross sections of the spheres, with the hollow space relatively small.

Required: How could we determine which of the two spheres is the hollow and which the solid one? There are no physical or chemical tools at your disposal. The spheres are not to be damaged, no holes drilled, no pieces cut out, and no paint removed. Try to think of a method so simple that we can decide this question in this room without any apparatus.

7. *The Pendulum Problem* (Maier, **17**)

Given: The subject is presented with four poles varying in length, one table clamp (C clamp), two burette clamps, two pieces of wire, several pieces of chalk, and several weights (pieces of lead tubing at Berlin, bolts at Michigan), and there are two chalk marks on the floor.

One pole must be at least as long as the distance between the chalk marks. Two of the remaining poles must together equal the distance from floor to ceiling, plus enough overlap for the C clamp to hold them.

Maier says, "All the experience necessary for the solution was given to the subject and did not have to be recalled by him" (**17**, p. 116).

Required: Make two pendulums, each holding a piece of chalk, which will swing over the chalk marks already on the floor, and will mark the floor.

This task was also used by Weaver and Madden (**37**).

8. *Blowing Out the Candles* (Maier, **17**, **19**)

Given: Among many objects lying on a table there are pieces of glass and rubber tubing, 6 to 12 in. in length, poles several feet in length, paper clamps; lighted candles are on a table some feet distant.

Hint 1. Water is siphoned from one container to another with a tube constructed like the one involved here. This is to give the subject the idea of the construction of the tube.

Hint 2. A piece of cord is hooked over a nail by the use of a pole. This is to bring the pole to the attention of the subject.

Required: Blow out the candles while standing several feet away (8 ft. in the later article, **19**).

This task was also used by Saugstad (**29**).

9. *The Two String Problem* (Maier, **18**)

Given: Two strings dangle from the ceiling too far apart to permit anyone to reach both of them at the same time, or to carry one over to where he can reach the other. There are many other objects in the room including a chair, a stick, more string, pliers, etc.

Hint 1. The experimenter brushed against a cord and set it swinging.

Hint 2. The subject was handed the pliers and told to use them.

Required: Tie the two pieces of string together.

This task was also used by Battersby, Teuber, and Bender (**2**), Birch and Rabinowitz (**4**), Cofer (**7**), Cofer and Gelfand (**8**), and Guetzkow (**12**).

10. *The Hat Rack* (Maier, **19**, **21**)

Given: The subject is presented with two poles, neither of which will reach the ceiling, but each more than half the required length, and a C clamp.

Hint 1. The two string problem is presented first, each string hung from an inverted **U**-shaped construction which utilizes the principle of the hat rack for each support pole. This is built by the experimenter, and left standing during the solving of the hat rack task.

Hint 2. The structure of Hint 1 is built and then removed before the hat rack task is presented.

Hint 3. The subjects who had been in the group in which the two string structure had been removed were asked to draw it.

(These hints were used in the later experiment, **21**.)

Required: Build a hat rack sturdy enough to hold a hat and a heavy coat.

This task was also used by Cofer and Gelfand (**8**).

11. *The Children's Swastika Pathway* (Maier, **20**)

Given: The child learns his way around a swastika-shaped path, with a booth at each end, the booths being different in size, shape, and furniture. The child then is shown a toy in one booth, the toy requiring a penny to make it work.

Required: The child is given a penny somewhere outside the apparatus, introduced into a booth other than the one with the toy, and expected to find his way to the toy.

12. *The Box Problem* (Duncker, **10**; and Adamson, **1**)

Given: On a table, among many other objects, lie a few tacks; three little pasteboard boxes (more or less the size of match boxes), apparently not identical in size, shape, color, position; and three candles. In one experimental situation, the candles are placed in one box, the tacks in another, and matches in the third.

Required: Put the candles side by side on the door of the room at eye height ("for visual experiments").

13. *The Gimlet Problem* (Duncker, **10**; and Adamson, **1**)

Given: Among other objects on a table lie three cords, two screw hooks, and a gimlet. In one experimental condition, there are three holes in a board on the wall, in the other the holes have not been bored.

Required: Hang the three cords from the board.

14. *The Paper Clip Problem* (Duncker, **10**; and Adamson, **1**)

Given: Many objects lie on a table, among them a piece of white cardboard, four smaller pieces of black cardboard, paper clips, glue or thumb tacks. A screw eye is fixed in the ceiling. In one condition of the experiment the black pieces are fastened to the white one with paper clips, in the other they are glued on (Adamson had them tacked on).

Required: Hang the cardboard arrangement from the ceiling screw eye.

15. *The Cut Pyramid* (Katona, **13**, Duncker, **10**; and Reid, **28**)

Given: An orthogonal tetrahedron (a four-sided pyramid, the sides being triangles of equal size), cut into two pieces along a plane defined by a line running between the midpoints of the sides of a triangle and parallel to its base, and a similar line parallel to the same edge of the tetrahedron and in another triangle.

Aid 1. What does this problem, to form a solid having only triangle surfaces, really mean?

Aid 2. Doesn't it mean forming large triangular surfaces, or at least one for a start?

Aid 3. Which two of all these surfaces when seen as extending together in the same plane, i.e., together forming one plane, ought to make that plane a large triangular surface, just the kind we are looking for?

(These aids were used by Reid, **28.**)

Required: Put these two pieces together to make a pyramid. (Duncker did not permit actual manipulation of the pieces, but had the subject tell him how to do it.)

16. *Breaking the Beads* (Bulbrook, **5**)

Given: The subject is presented with a string of beads arranged to alternate one yellow with two white, but with five white beads in the center. Other objects (pliers, needle, saw, etc.) lie on the table.

Required: Instruction 1. Make a single regularly repeated pattern without either unstringing or restringing the beads, and without either knotting or breaking the thread. Proceed aloud.

Instruction 2 was like the above except that, "Use any of the means supplied on the table" was inserted between the two sentences.

Note: The task following the one above in Bulbrook's article was similar to it. The string was made of unpainted wooden beads and ink stained wooden beads and the subject had to stain the extra beads with ink from a bottle on the table.

17. *The Balloon in the Bottle* (Bulbrook, **5**)

Given: A deflated toy balloon, a 75-cc. flask, a penstaff, pencils, a file, were presented to the subject. In group experiments the experimenter put the balloon into the flask and blew into the balloon.

Required: To blow up the balloon inside the bottle.

18. *Weighing the Coins* (Simmel, **31**)

Given: The subject is to imagine he has eight coins, which may or may not include one counterfeit which would be lighter than the others. He also has an imaginary balance.

Required: In two weighings, identify the lighter coin or prove that there is none.

Simmel also used nine coins, one certainly counterfeit, in two weighings; twenty-five coins, one certainly counterfeit, in three weighings.

19. *The Shooting of the Captain and His Son* (Burack, **6**)

Given: The subject is presented with a printed problem referring to two men shot to death at close range, the gun being under one body, no footprints except their own, and so on.

Required: The subject is asked to tell whether each man was killed by accident, suicide, murder, and later is asked more specific questions such as, "Was there information which would tell which man died first?" etc.

20. *The Cylinder in the Can* (Saugstad, **29**)

Given: A wooden cylinder is placed upright in a can into which it fits loosely. Apparently the top of the cylinder should be a bit below the top of the can, and the fit should be close enough that one cannot put his fingers in between the can and the cylinder. On the table beside the can are a penholder, pliers, thread, screw driver, pitcher of water, drinking glasses.

Required: Get the cylinder out of the can without tipping the can.

21. *Twenty Questions* (Lindley, **16**; Taylor and Faust, **34**)

Given: The subject is told that the experimenter is thinking of an object and whether it is in the animal, vegetable, or mineral kingdom. The subject is to ask questions which narrow the field of possibilities, each question being answered by "Yes," "No," "Partly," "Sometimes," or "Not in the usual sense of the word." A question which cannot be so answered must be restated. This description applies to Taylor and Faust's procedure.

Required: The subject is to be able to name the object with thirty questions or less (Taylor and Faust).

22. *The Incorrect Square Root* (Marks, **22**)

Given: The subject is taught a method of finding square roots using a calculating machine and a table of factors. By means of incorrect entries in the table, the subject gets wrong answers to two out of four problems.

Required: Ascertain the source of the error.

23. *The Disc Transfer* (J. Peterson and Lanier, **25**)

Given: The subject is given three stationary circles (e.g., painted on a board) and several discs varying in diameter, stacked on one circle in order of size, the smallest at the top. (If one were going to make himself a set of discs, it would be easier to make squares.)

Required: The subject is to transfer the stack to another circle, moving one disc at a time, using the third circle as a way station, never putting a disc on a smaller one. He starts with two discs, then three, and so on to seven.

This task was also used by Shaw (**30**).

24. *Making the Last Draw* (J. C. Peterson, **26**)

Given: The subject is presented with a set of seven beads on a string (or seven matches or pennies or other objects lying on a table). He is told that he and the experimenter will draw from the set alternately, each taking one or two beads as he wishes. The subject gets the first turn and is to try to draw the last bead.

Required: To learn to win the game every time with different numbers of beads; to find a general principle for winning (*a*) with draws of one or two beads, (*b*) with draws of other numbers.

This task was also used by Waters (**36**).

25. *The Electromaze* (Kruglak, **15**)

Given: The apparatus presents the subject with four push buttons arranged in a square, with a fifth button in the middle. On top of the box are a red light and a green light. The center one of the five buttons is an "erase" button.

Required: The subject is to solve a series of seven problems, most of which are sequences of button pushes. The red light indicates an error in the sequence and the subject then pushes the "erase" button and starts again on a new trial. The green light appears only at the end of a correct sequence.

26. *Water Jars Problems* (Terman, **35**)

Given: Typically, the subject is told to imagine that he has two or three containers of specified sizes, although sometimes actual containers are used.

Required: To tell the experimenter how to use the containers to secure a given amount of water, no estimations (as of "half") being permitted.

This sort of task has been used by many people, lately especially in investigations of set.

27. *The Multiple-Choice Apparatus* (Yerkes, **39**; and Beasley, **3**)

Given: In the form of the equipment used for humans, the subject sits before an apparatus from which any one or any set of twelve keys can be protruded toward him. Repeated problems can be presented by different sets of three keys each, or by successive sets of keys which vary in number except that the number is always odd, and so on.

Required: To find the key which will cause a buzzer to sound; to find a rule by which to identify the correct key in future sets.

Note: This task has a correlative form which can be used with animals, and it is ordinarily assumed that the problem is the same for human and animal subjects.

28. *McGeoch's Problem Box* (Ericksen, **11**)

Given: The subject is given a box, 8×8×12 in., with three knobs protruding from each of two opposite sides, and a seventh knob on one end. The knobs can be pushed, pulled, turned, or ignored, making a complex trial and error problem.

Required: Remove the end containing the knob.

29. *Moments of Levers* (G. M. Peterson, **24**)

Given: A lever is set on a table with hooks at two-inch intervals out from the fulcrum, the hooks being plainly numbered 1, 2, 3, 4 from the center out, on each side. Ten weights, all of the same diameter and of different lengths, are provided, again plainly numbered 1, 2, 3, 4, 5, 6, 8, 9, 10, 12.

Required: The experimenter says, for example, "If you had a 1 ounce weight on 4 (he puts it there), where would you put a 2 ounce weight to balance it?" The subject gives his answer in writing and the weight is not actually hung on the lever to show that it does or does not balance. The subject receives no knowledge of results, neither by actual trial, nor by any statement from the experimenter. Instead, each problem is followed by a demonstration of a different problem. There are twenty problems and twenty demonstrations. The subject is now asked to state the principle involved.

ATTRIBUTES WHICH MAY BE USEFUL IN PROBLEM-SOLVING TASKS

It is possible to offer a set of attributes which may be found useful in problem-solving tasks, although none of the tasks considered in the paper meet the complete list. In fact, it may be doubted that any one task ever will exhibit all of them, and some of them are certainly unsuited to certain kinds of investigations, as for example, "only one way of solving the problem" is certainly not desired in most of the current experiments on set. The attributes are presented as suggestions to be considered in choosing or designing tasks, and in the hope that they will serve as a base from which improved lists of criteria can be constructed.

Two important criteria for problem-solving tasks are that the task should provide a scoring continuum longer and more useful than the two-step continuum of pass-fail scores which have been frequent in problem-solving tasks, and that it should provide the experimenter with a maximum amount of knowledge of what the subject is doing. Hypotheses must be regarded as intervening variables, definable on the basis of the task presented and of the overt responses produced. The experimenter's knowledge of what the subject is doing is provided by his knowledge of the task and by these overt responses. Other more detailed characteristics of tasks may be listed under those headings, some of them falling under both.

Score Continuum

A continuum of scores from a task will be useful in differentiating among subjects and among experimental groups, thus increasing the possibility of finding functional relationships with independent variables. The scores should all lie along the same dimension—a matter of knowing what the subject is doing, or, in other terms, of factorial purity and of validity (see below).

A continuum may be found in (*a*) the number of unsuccessful trials preceding solution as in Kruglak's Electromaze (task 25) where the subject has to discover a sequence by trial and error and one can count the number of errors; (*b*) the number of problems from a series on which the subject was successful, as, again, in the Electromaze where several sequences are to be discovered, or as in Water Jars problems (task 26) where one can use as many problems as he wishes, or in the Children's Swastika Pathway (task 11) where the child is asked to find the goal from different points within the swastika, or in Burack's Shooting of the Captain and His Son (task 19) where there is a series of questions to be answered, or in G. M. Peterson's Moments of Levers (task 29) where 20 problems are given with the same lever, but with differing weights and distances, or in the Multiple-Choice Apparatus (task 27); (*c*) the number of steps on which the subject succeeded (see paragraph below); and (*d*) the amount of assistance provided before solution as was done by Maier with his Pendulum Problem (task 7) and by Reid (**28**) in an experiment with Katona's Cut Pyramid (task 15). Deutsche (**9**) turns a qualitative answer into a quantitative one by providing an eight-point scale against which to compare it (see footnote 2). Other possibilities for providing continua are given below under Types of Scores.

Time to solution always provides a possible scoring continuum and has been used frequently, but it is possible that time is not always a desirable measure of problem-solving performance, because, for example, the events which occupy that time may not all lie on the same behavior dimension. Consider also Maier's statement that "some subjects worked as long as three hours, those who got the solutions worked five or ten minutes" (**17**, p. 121).

Subgoals

One possible method of providing a continuum of scores is to use a problem which includes several steps, several subgoals on the way to the final solution. This is to be differentiated from a series of problems as that phrase was used in the preceding section, in that in a series of problems each is independent of the others, whereas in a series of steps, what the subject discovers on one step is, ideally, used on the next step. A nice example of a problem containing several steps is provided by Making the Last Draw (task 24), where the subject learns to solve a problem with successively larger numbers of manipulanda and where he typically makes his generalizations about the method more and more inclusive as he goes along. Another example is given in Simmel's Weighing the Coins (task 18), where the subject has to identify the odd coin from among eight, then from among nine, then from twenty-five. Peterson and Lanier's Disc Transfer (task 23) is another series of steps in which increasing numbers of pieces can be used. Increasingly complex tasks can be presented with Yerkes' Multiple-Choice Apparatus (task 27).

It is probably much too early in the history of research in problem solving to ask for a set of steps such that if a subject fails on one step we can be sure that he will fail all succeeding steps, but this would be an advantageous feature of any complex task, especially if the task were ever to be used in measuring problem-solving abilities or aptitudes. If the steps of the problem cannot be arranged in any such scale as that implied in the previous sentence, then the task must provide every subject with an opportunity to try every step. An example of a task where the latter does not happen is Maier's Pendulum Problem (task 7), where the subject must construct a pendulum which will make marks across particular spots on the floor and where he can hardly begin on the marking step of his problem until he has constructed the pendulum.

Few Hypotheses

It will assist the experimenter in knowing what the subject is doing if the possibilities for action are limited, which is to say that the task should permit only a small number of hypotheses, including both apposite and inapposite hypotheses. The number of possible hypotheses is different from the number of steps in a problem or the number of problems in a series since the number of hypotheses includes the possible inapposite hypotheses which the subject may propose. Maier's Blowing Out the Candles (task 8) requires two steps in the construction of the tube, the joining of the pieces and the stiffening of the construction, each with its accompanying hypothesis, and an additional step and hypothesis in blowing through the tube. Maier also, however, lists additional inapposite hypotheses such as pouring water down the tube; so the number of hypotheses is greater than the number of steps. The opposite sort of example, of an indeterminate number of hypotheses, is provided by Twenty Questions (task 21) and by Yerkes' Multiple-Choice Apparatus (task 27). Insofar as most of the tasks being considered here are one–apposite-hypothesis kinds of problems, it is easy to list their apposite hypotheses as in Duncker's Box Problem (task 12), where the only hypothesis needed is that the box can be fastened to the wall to make a shelf, or in Bulbrook's Breaking the Beads (task 16), where the idea of smashing the beads is the only idea needed. It is rare, however, that an investigator attempts to list all of both the apposite and inapposite hypotheses, although one man does so when Marks' description of the Incorrect Square Root Problem (task 22) says that the difficulty may lie in the operator, the calculating machine, the method, or in the table of factors. Even here, however, one may wonder whether subjects cannot produce further hypotheses, as, for instance, in producing several about the particular fault which may be occurring in the operator.

One Solution

Having only one solution, and

only one method of arriving at that solution for a problem, or for the steps or parts of the problem when there are any, will contribute further to the experimenter's knowledge of exactly what the subject is doing. These attributes will also contribute to factorial purity and to reliability. Maier's Two String Problem (task 9) may be mentioned here to emphasize the statement made in the introduction to this section of this paper where it was pointed out that not all of these attributes would be desirable in all problems. The Two String Problem has four or five possible solutions[3] and that feature of the task is an essential feature of, for example, one of Guetzkow's experiments (**12**).

Multiple Scoring

More than one method of scoring the subject's behavior may sometimes bring surprises, as when two methods do not agree in their ratings of the subjects' performances. Such disagreements may lead to a finding that different dimensions or aspects of the behavior are being measured, or, of course, to the conclusion that one or all of the scores are unreliable. On the other hand, if the scoring methods do agree one will be justified in thinking them the more reliable. Several methods of scoring are presented below in the section on scoring. We have a fine example of ingenuity in the invention of scoring methods when Burack uses ten (task 19).

Overt Behavior

If the experimenter is to get a maximum amount of information about subject behavior, it will obviously be of advantage that as much as possible of the behavior be overt, with any covert behavior being treated as intervening variables. Such overt responses could be verbal if they are parts of, rather than reports of, the solving. Equipment problems are more likely to produce overt behavior than are other types, and ideally each possible apposite or inapposite hypothesis would be expressed in some manipulation of the equipment, different hypotheses producing different manipulations. There are tasks in which all of the apposite, and perhaps even all of the inapposite, hypotheses are expressed overtly. Bulbrook's Balloon in the Bottle (task 17) is a case in point. The subject must inflate the balloon inside the bottle, and to do so must insert into the neck something to prevent the balloon from completely filling it and thus preventing escape of the air. Inserting a pencil or a glass tube into the bottle neck would seem a fairly good indication that the subject is using the air escape hypothesis. If Simmel (task 18) had provided actual instead of imaginary balances and coins, she might have had an overt expression of an inapposite hypothesis when the subjects tried to solve the problem by dividing the coins equally between the two pans of the balance. Saugstad's Cylinder in the Can (task 20) offers opportunities to watch the subject manipulate a water pitcher as the result of a hypothesis about floating the cylinder high enough to permit grasping it and removing it from the can. Saugstad also provides an opportunity for inapposite hypotheses to exhibit themselves by overt, manipulatory behavior as when an attempt is made to loop a thread around the cylinder, to grasp it with pliers, to insert one's fingers into the can, etc. It may be, however (and the present writer has not seen the equipment), that these hypotheses could be rejected on the basis of visual cues, which produces the same practical results as if they were rejected on a completely covert basis. Kruglak's Electromaze (task 25) seems to offer a fine opportunity to see subjects adopt hypotheses although Kruglak makes no reports of such aspects of their work. These examples suggest that it may be possible to design tasks that will do excellent jobs of turning internal, symbolic behavior, into external, observable behavior.

---

[3] "Four or five" because although Guetzkow and Maier list four, Duncker (**10**, p. 23) lists a fifth, which, however, does not seem to me to be a solution.

Unidimensionality

Another characteristic of problem-solving tasks which would increase the experimenter's knowledge of what the subject is doing would be to have him doing only one kind of thing at a time; for example, using only one hypothesis at any one moment, or, in other cases, using only one type of process (as deduction, induction, perceptual reorganization). This is a request for what the factor analysts call factorial purity and what the experimentalists call unidimensionality. It implies more knowledge than is presently available about such intervening variables as hypotheses and processes, and about tasks demanding overt responses which will differentiate one intervening variable from another. On such bases we could choose tasks which included only one dimension, or, when desired, tasks including more than one dimension, but in which the dimensions are known and differentiated by different kinds of responses. Work on this sort of thing is proceeding apace among the factor analysts, but there is very little of it in the more complex apparatus task studies.[4]

Granting that there is available little knowledge of intervening variables in problem solving, or of the operations which will discriminate among them, it is still possible to rule out patently "extra" abilities such as the arithmetic ability needed in a water jar problem which includes such numbers as 163, 29, 11; although it may be permissible to include them in cases where it is fairly certain that every member of the experimental population has the ability in a high degree, as when Duncker and Adamson expect all their subjects to be able to bore holes with a gimlet (task 13) and to bend paper clips (task 14), or when Marks expects his subjects to have enough arithmetic ability to learn to find square roots (task 22). Examples of the other kind of task, that which is multidimensional, are apparently provided by the Hollow and Solid Spheres (task 6) and by the Shooting of the Captain and His Son (task 19).

Special Knowledges

As in the case of extra abilities, so extra and special knowledges should be avoided or should be identified and when necessary, provided. Examples of this latter are Székely's tasks (tasks 1 to 6) in connection with which he sometimes asked the subjects for the required information, sometimes provided it, etc. In

---

[4] It is possible to find an example of an attempt to identify psychological dimensions in a problem-solving task by going outside the present realm of discourse. Cornell, in an unpublished Twelfth Quarterly Report on an Air Force contract, uses a "Tab Test" which is an analogue of a radar set, and is a printed test in which a trouble shooter can look for symptoms, make checks of components, etc. He finds that the behavior can be classified tentatively into such categories as efficiency, misinterpretation, precipitancy, caution. Whether or not these categories are eventually included in the final set, they are the general sort of thing that dimensions are—and they provide four different methods of scoring a task.

Marks' problem of the Incorrect Square Root (task 22), the subject has to be taught a special method for finding square roots before being able to work on the task at all. In the Pendulum Problem (task 7), Maier provided all the necessary "experience" and asked the subject to assemble it properly. Twenty Questions (task 21), on the other hand, is a case where the knowledge needed could never be specified. The Cylinder in the Can (task 20) gives an instance of information (that wood floats) which can be safely assumed to be within the repertoire of all of the members of many different groups of subjects. It has already been suggested (footnote 2) that perhaps some of Piaget's tasks (**27**) are tests of special knowledge rather than of reasoning.

Reliability

Problem-solving tasks, like any other experimental tasks, should be reliable in the usual statistical, test-construction meaning of that word, which is to say that the variance between subjects should be produced by features of the task and/or of the independent variable under examination, rather than by chance factors such as events during the period immediately preceding entrance to the experimental room. This matter of reliability is one which has never been solved very satisfactorily for experimental tasks, although the need has been apparent for many years (see Melton, **23**), and is an exceptionally baffling question in problem solving where the subject frequently meets a task only once. Signs of an interest in reliability among the tasks described above appear (*a*) in connection with the Disc Transfer Problem (task 23), where correlations in the region of .10 are reported between scores on the problem with four discs and with five discs; (*b*) in G. M. Peterson's reported split-half reliability of .95 with twenty highly similar tasks all involving the moments of a lever (task 29); and (*c*) in Deutsche's work (footnote 2). Unreliability of tasks will presumably not lead to unwarranted generalizations, but may result in failure to find relationships which would otherwise appear.

Validity

If by validity one refers to the question as to whether or not any one task really demands problem-solving behavior, the answer will lie in expert judgment and in definitions which can be described as expert judgment again. Identification of types of processes which occur in what experts agree to be problem solving will, in turn, add to the assurance of validity in future selection of tasks. As far as the individual experimenter is concerned, he can frequently avoid the question of validity by refusing to label his task as problem solving, describing it completely enough that anyone interested in the matter can make his own judgment. A more important question than the above is whether the behavior in question has any relation to any other kind of behavior, which brings us to the other definition of validity: the relationship of performance on one task to performance on another, for instance, the relation between performance on a test and performance on an equipment trouble-shooting job. The split-half reliability of a series of tasks is this sort of validity (task 29 and footnote 2).

The reader is again referred to Melton (**23**) for a more extended discussion of this topic.

TYPES OF SCORES USABLE IN PROBLEM-SOLVING TASKS

It is possible to make a list of various ways of scoring problem-solving tasks, developing the list from the methods used in conjunction with the tasks themselves, from remarks made by other persons in other places, and from reflection on these sources. The simpler kinds of scores require no examples and some of the more complex ones have none.

Quantitative

Easily quantifiable

1. Success or failure
    - *a.* On the problem as a whole
    - *b.* On steps in a problem
    - *c.* On each of a series of problems
2. Error scores
    - *a.* Omits correct procedures, hypotheses, etc.
    - *b.* Uses correct procedures in the wrong places
    - *c.* Repeats correct procedures unnecessarily
    - *d.* Adds incorrect procedures
3. Time scores

Not as easily quantifiable

1. Assistance scores. This is a question of how many aids or hints were required to produce the solution and is therefore a sort of success or failure score. This sort of score was used by Maier, for instance with the Two String Problem (task 9) and with the Hat Rack (task 10), and was used by Reid (**28**) with Katona's Cut Pyramid (task 15).

2. Deficiency scores. How close to solution did the subject come? How many and how large were the steps yet to be taken?

3. Elegance and neatness of the solution. If there are two methods of solving the task does the subject take the shorter or easier or the one with the higher probability of success? This is the sort of scoring done in Twenty Questions (task 21) and is suggested in Burack's asking "Was the problem clearly formulated by the subject?" (task 19). Deutsche's (footnote 2) use of a scale against which to compare written answers is an example of quantifying the elegance of "solutions."

Qualitative

1. Type of psychological process used in the solving; as, deduction, perceptual reorganization.

2. Content of the process; what deduction or what hypothesis does the subject use? Simmel does this in her experiment on Weighing the Coins (task 18) when she looks to see whether the subject is reacting on the basis of symmetry (having the same number of coins on each pan of the balance), totality (weighing all the coins simultaneously), or divisibility (dividing the coins into groups which will leave no extra, nongrouped coins).

3. Elegance of the solution process. Compare Wertheimer (**38**, p. 18) *et al.* on sensible procedures which evince some understanding of what is to be done, and blind procedures such as trying to apply a formula without first deciding whether the formula fits.

4. Deficiency scores. What necessary steps were omitted? There is a use of this sort of thing when Burack (The Shooting of the Captain and His Son, task 19) remarks (6, p. 12) "Her solution would be satisfactory except for ignoring. . ."

5. The nature of the error. This is the opposite of elegance of the solution process. Compare Duncker's (**10**, p. 6) meaningful and nonsensical errors and Köhler's (**14**, p. 203) good, uninformed, and stupid errors.

6. Content of the error; what specific errors appeared? There are many examples of these errors in the articles describing our tasks as when one subject wanted to put Bulbrook's

(task 17) balloon clear into the bottle and blow it up with a glass tube—except that there was no tube available. This would be a "good" error in the preceding category.

DISCUSSION

As one reads through the experimental literature of complex problem solving as defined by the sort of tasks included here, he is struck with the fact that many of these tasks have been of the parlor puzzle or game variety rather than being tasks constructed for experimental use, which is even more true in the problems which have been omitted than in those included. Such problems are frequently either too simple in that the solving process is a single, rapid response (and often a covert response), or too complex in that there are so many possible responses that the experimenter has little chance of identifying them. This, however, is as often true in the tasks designed for experiments as it is in those taken over from the parlor. The choice of such problems is presumably due to the many investigations made in this field which have been observational rather than experimental in their aim, although this is not so true today as it was in the past.

Few of the tasks described herein permit more than a two-category scoring system, success or failure, although some investigators have shown considerable ingenuity here. Reliability has been mentioned only twice and validity never by the writers of the reports. Only seldom are all the cogitable apposite hypotheses listed, and cogitated inapposite hypotheses are sometimes stated, but a complete list of them for any one task has apparently never been offered.

There is a reiterated plea, especially in articles and chapters which summarize the state of the field of problem solving and reasoning, for more theory—a plea which has produced no great results. Perhaps a more modest sounding plea may be made. What this field needs are dimensionable independent variables, predictions of their effects on dimensionable dependent variables, and tasks especially designed to measure those effects. Once this is achieved, we can start to work on hypothetical constructs, perhaps using specific hypotheses as first order intervening variables and processes such as hypothesis formation by induction or deduction as second order variables.

From one point of view, research in problem solving is simpler than research in learning, in that in learning the experimenter predicts changes in responses, whereas in problem solving all he has to predict is the probability of occurrence of the response. From another point of view, however, it must be admitted that research in problem solving may be more difficult because a greater number of more complex, intervening variables may be needed here than in learning research.

REFERENCES

1. ADAMSON, R. E. Functional fixedness as related to problem solving: a repetition of three experiments. *J. exp. Psychol.*, 1952, **44**, 288–291.
2. BATTERSBY, W. S., TEUBER, H. L., & BENDER, M. B. Problem solving behavior in men with frontal or occipital brain injuries. *J. Psychol.*, 1953, **35**, 329–351.
3. BEASLEY, W. A miniature model of Yerkes' multiple-choice apparatus. *Amer. J. Psychol.*, 1931, **43**, 497–498.
4. BIRCH, H. G., & RABINOWITZ, H. S. The negative effect of previous experience on productive thinking. *J. exp. Psychol.*, 1951, **41**, 121–125.
5. BULBROOK, MARY E. An experimental inquiry into the existence and nature of "insight." *Amer. J. Psychol.*, 1932, **44**, 409–453.

6. BURACK, B. The nature and efficacy of methods of attack on reasoning problems. *Psychol. Monogr.*, 1950, **64**, No. 7 (Whole No. 313).

7. COFER, C. N. Verbal behavior in relation to reasoning and values. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie, 1951.

8. COFER, C. N., & GELFAND, S. The role of increased associative strength in reasoning: further report. Twelfth Technical Report under Contract N7-onr-397, T. O. III, Univer. of Maryland, 1952. Duplicated.

9. DEUTSCHE, JEAN M. *The development of children's concepts of causal relations.* Minneapolis: Univer. of Minnesota Press, 1937.

10. DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).

11. ERICKSEN, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, **31**, 339–345.

12. GUETZKOW, H. An analysis of the operation of set in problem solving behavior. *J. gen. Psychol.*, 1951, **45**, 219–244.

13. KATONA, G. Eine kleine Anschauungsaufgabe. *Psychol. Forsch.*, 1927, **9**, 159–162.

14. KÖHLER, W. *Mentality of apes.* New York: Harcourt, Brace, 1925.

15. KRUGLAK, H., SCHENSTED, C. E., & SELF, H. C. An electric multiple choice maze. Technical Report Number 2 under Contract N8onr-66213, Project NR192-041, Univer. of Minnesota, 1952. Duplicated.

16. LINDLEY, E. H. A study of puzzles with special reference to the psychology of mental adaptation. *Amer. J. Psychol.*, 1897, **8**, 431–493.

17. MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.

18. MAIER, N. R. F. Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *J. comp. Psychol.*, 1931, **12**, 181–194.

19. MAIER, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, **24**, 144–155.

20. MAIER, N. R. F. Reasoning in children. *J. comp. Psychol.*, 1936, **21**, 357–366.

21. MAIER, N. R. F. Reasoning in humans: III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.

22. MARKS, M. R. Problem solving as a function of the situation. *J. exp. Psychol.*, 1951, **41**, 74–80.

23. MELTON, A. W. The methodology of experimental studies of human learning and retention: I. The functions of a methodology and the available criteria for evaluating different experimental methods. *Psychol. Bull.*, 1936, **33**, 305–394.

24. PETERSON, G. M. An empirical study of the ability to generalize. *J. gen. Psychol.*, 1932, **6**, 90–114.

25. PETERSON, J., & LANIER, L. H. Studies in the comparative abilities of whites and negroes. *Ment. Measmt Monogr.*, 1929, Serial No. 5.

26. PETERSON, J. C. The higher mental processes in learning. *Psychol. Monogr.*, 1920, **28**, No. 7 (Whole No. 129).

27. PIAGET, J. *The child's conception of physical causality.* New York: Harcourt, Brace, 1930.

28. REID, J. W. An experimental study of "analysis of the goal" in problem solving. *J. gen. Psychol.*, 1951, **44**, 51–69.

29. SAUGSTAD, P. Incidental memory and problem solving. *Psychol. Rev.*, 1952, **59**, 221–226.

30. SHAW, MARJORIE E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, **44**, 491–504.

31. SIMMEL, MARIANNE L. The coin problem: a study in thinking. *Amer. J. Psychol.*, 1953, **66**, 229–241.

32. SZÉKELY, L. Knowledge and thinking. *Acta Psychol.*, 1950, **7**, 1–24.

33. SZÉKELY, L. Productive processes in learning and thinking. *Acta Psychol.*, 1950, **7**, 388–407.

34. TAYLOR, D. W., & FAUST, W. L. Twenty questions: efficiency in problem solving as a function of size of group. *J. exp. Psychol.*, 1952, **44**, 360–368.

35. TERMAN, L. M. Genius and stupidity. *Ped. Sem.*, 1906, **13**, 307–373.

36. WATERS, R. H. The influence of tuition upon ideational learning. *J. gen. Psychol.*, 1928, **1**, 534–549.

37. WEAVER, H. E., & MADDEN, E. H. "Direction" in problem solving. *J. Psychol.*, 1949, **27**, 331–345.

38. WERTHEIMER, M. *Productive thinking.* New York: Harper, 1945.

39. YERKES, R. M. A new method of studying the ideational behavior of mentally defective and deranged as compared with normal individuals. *J. comp. Psychol.*, 1921, **1**, 369–394.

Received March 14, 1954.

Psychological Bulletin
Vol. 52, No. 2, 1955

SPECIAL REVIEW

RECENT BOOKS IN CHILD PSYCHOLOGY

BOYD R. McCANDLESS
Iowa Child Welfare Research Station

Baker, Harry J. *Introduction to exceptional children.* New York: Macmillan, 1953. Pp. xvi+500. $5.00.

Crow, Lester D., & Crow, Alice. *Child psychology.* New York: Barnes and Noble (College Outline Series), 1953. Pp. xvi+267. $1.50.

Cruze, Wendell W. *Adolescent psychology and development.* New York: Ronald Press, 1953. Pp. xii +557. $5.00.

Davidson, Audrey, & Fay, Judith. *Phantasy in childhood.* New York: Philosophical Library, 1953. Pp. viii+188. $4.75.

Ganz, Madelaine. *The psychology of Alfred Adler and the development of the child.* (Trans. by Philip Mairet.) New York: Humanities Press, 1953. Pp. xxiv+203. $4.50.

Halpern, Florence. *A clinical approach to children's Rorschachs.* New York: Grune and Stratton, 1953. Pp. xvi+270. $6.00.

Hurlock, Elizabeth B. *Developmental psychology.* New York: McGraw-Hill, 1953. Pp. x+556. $6.00.

Jersild, Arthur T. *Child psychology.* (4th Ed.) New York: Prentice-Hall, 1954. Pp. xii+676. $6.00.

Martin, William E., & Stendler, Celia Burns. *Child development: the process of growing up in society.* New York: Harcourt, Brace, 1953. Pp. xxiv+519. $4.75.

Moustakas, Clark E. *Children in play therapy: a key to understanding normal and disturbed emotions.* New York: McGraw-Hill, 1953. Pp. x+218. $3.50.

Peck, Leigh. *Child psychology: a dynamic approach.* Boston: D. C. Heath, 1953. Pp. vi+536. $5.25.

Stolz, Lois Meek, & Collaborators. *Father relations of war-born children.* Stanford, Calif.: Stanford Univer. Press, 1954. Pp. viii+365. $4.00, paper.

Wolffheim, Nelly. *Psychology in the nursery school.* (Trans. by Charles L. Hannan.) New York: Philosophical Library, 1953. Pp. 144. $3.75.

Zubek, John P., & Solberg, Patricia Anne. *Human development.* New York: McGraw-Hill, 1954. Pp. viii+476. $6.00.

Introduction

A special review is presumed to represent, among other things, the reviewer's vectorial description of the field for the period covered by the review; i.e., where is the field moving? This review covers the 14 books (12 hardbacks, two paperbacks) available to the reviewer. They fall into five classes:

First, child and developmental psychology, four volumes (Crow and Crow, Jersild, Martin and Stendler, and Peck). Two volumes of human development (development during the entire age span, and for Zubek and Solberg, the phylogenetic series) are also included here (Hurlock, and Zubek and Solberg). There is a single volume on adolescence (Cruze). All seven books are designed as texts.

A second class has to do with exceptional children, one volume (Baker) which is also intended as a text.

Third, three volumes can be classified as being clinical or other applications of child psychology. Two of the volumes deal with children's fantasy (Davidson and Fay, and Moustakas) and one with their Rorschachs (Halpern). In the sense of a distinction made below, all can be considered child psychology.

The fourth class, one volume, is a research monograph (Stolz and coworkers) and would be classed as child psychology. The fifth class, two volumes, might be labeled psychoeducation (Ganz and Wolffheim).

These 14 books encompass about 5,900 pages and do not include the newest revision of Carmichael (**2**). In one sense, all are psychological in intent. Most of the American authors are listed in the APA directory, and the purpose of each book seems to be to present a picture (varying in completeness) of children's behavior development with the goal of guiding it in the direction of a better child and adult adjustment. It is true that physical growth, anthropological, and sociological material is included in all the texts, but the ultimate intention of such inclusion is to increase the understanding and control of behavior.

The reviewer would like to present a tentative distinction between child development and child psychology. It is possible to study the child from many points of view (e.g., nutritionally, in terms of growth, anthropologically, psychologically, sociologically). A child *development* text draws from several or all of these fields, although most child development personnel are classified professionally as psychologists. Strictly speaking, a child *psychology* text draws only from the psychological area. Historically, child development studies and texts have been practically and pragmatically oriented. Their goals are those implicit in the volumes in classes 1 and 2. In this sense, all the volumes in classes 1 and 2 are child development, not child psychology texts.

In the same sense, the three volumes in class 3 and the research monograph in class 4 are child psychology since they draw from traditionally psychological materials and methods (e.g., fantasy, projective techniques, the influence of certain learning conditions on personality development). Their emphasis is more that of understanding than of controlling behavior.

There are two types of intellectual affiliation typical of child psychologists: the traditional child development affiliation, and the general psychology affiliation. The first is desirable for practical reasons, if no other than that it has been the historical pattern: in this field, in their teaching, service, and research functions, child psychologists mingle with many disciplines and are confronted with many practical questions; the average research center using children as subjects is multidisciplinary, etc. The second affiliation is equally desirable but is less characteristic than the first of professional personnel in the field. The goal of the science of psychology is to understand behavior and the goal of these volumes and of child psychologists in general is to understand the behavior of children. A divorce between general and child psychology is self-defeating for all concerned.

There are certain indications that the authors of these volumes (and child psychologists in general) have not even entered into a marriage with general psychology, let alone being at the divorce stage. Their preoccupation appears to be with the child as a special subject of study in his own right; it is not with the child as an organism whose behavior, studied systematically, will add scope and precision to the subject matter of the science of psychology. These goals are not mutually exclusive by any means; rather they are richly supplementary. Professionally, neither psychology nor child psychology (a merely practical distinction) can afford not to strive toward both of them.

Illustrative of this criticism of these 14 books is the fact that most of the material in them is pragmatic and nonsystematic. This is less a criticism, of course, of the authors than of the profession, past and present. Nevertheless, there is a tendency to omit from the texts some of the more important theoretically based psychological studies. For example, four studies (**1**, **4**, **5**, **6**) were not reported by any of the authors although the datelines indicate that they were available in the literature at the time all the textbooks were being written. Is this because three of the four were published in the *Journal of Experimental Psychology*? Is it because all four articles are systematic and theoretical rather than descriptive in intent? Surely the subject matter of the studies (discrimination learning and transposition, secondary reinforcement and perceptual learning, and compensatory behavior as a function of implied parent-child relations) is important.

With the exception of the Martin and Stendler text, theoretical matters are treated scantily if at all; the same is true, with the same exception, of sociological and anthropological literature. Only Jersild adopts an ordering frame of reference (the self theory as represented by Sullivan and others).

The appearance of two volumes dealing with over-all human development is new for the field. If such greater span of coverage results in perspective and increased systematization, the addition seems desirable. If it results only in increased, encyclopedic cataloguing of normative information without coherent frames of reference, no advantages accrue.

Encouragingly, the texts as a group pay more attention to methodology and questions of research adequacy than has been true traditionally, although there are some exceptions to this tendency.

It is difficult to say anything as a class about the special clinical and application volumes. They are variable in quality and their principal value seems to be technical and practical. Some intriguing research possibilities are opened up by the psychoanalytic educational paradigms suggested by Ganz and Wolffheim.

Most of the volumes reviewed here can be criticized in spots for one or more of the following common errors. An attempt has been made to classify and illustrate these errors, which are frequent not merely among the authors of the present volumes but for psychology authors in general. Some of such flaws, of course, are inherent in the process of compiling material and making it comprehensible.

A first flaw is poor analysis, in terms of the philosophy of science. Illustratively: "Whenever young children are placed in favorable environments results of intelligence tests tend to rise, but it is very doubtful whether this is a real rise in mental level or a spurious change due to the nature of testing materials" (Baker, p. 238).

Martin and Stendler (p. 193) ask analogous questions re IQ changes: "Are such changes *real*? Do they really represent changes in the intellectual level of the children involved?"

With reference to these quotes, and others similar are present in most of the texts, we might say that if an increase in intelligence test scores is accompanied by increases in other performances with which it (intelligence test score) is positively correlated, then in the only meaningful sense of the term "real," the changes are real.

A second type of error consists of overgeneralization from available data. Illustratively: Hurlock (p. 13) states, " . . . learning must wait upon maturity. Trying to teach a child something before he is ready will invariably lead to failure. This will be accompanied by a resistant attitude on the child's part to further learning in the same area or even in related areas. As a result the child will not learn what he is capable of learning when a state of readiness occurs."

This quote is selected because it reflects the considerable preoccupation with the issues of maturation-learning (and heredity-environment) common to most of the texts. The value issues around these questions are moderately handled in all cases; but most of the authors show some unclarity which should by now have been dissipated by familiarity with the philosophy of science. This is to say: with our present methods (which are determined by the essential inviolacy of human unions and the human organism) we cannot construct designs which make it possible to answer the relevant questions empirically (e.g., one cannot put one identical twin in isolation, the other in a maximally stimulating environment; babies up for adoption *are* selectively placed; selective mating is not possible; training to the extent of possible damage to the child is socially and legally impossible). Zubek and Solberg, in analogies from animal research, deal most sophisticatedly with such issues.

Through the texts runs the thread of the classical maturation-learning studies (e.g., the Gesell and McGraw twins). The conclusions from these that the effects of early training are minimal are accepted by most of the authors, as they are by Hurlock. Other lines of evidence are not considered: e.g., the Fels studies suggest profound personality and intellectual effects resulting from parental accelerationism; scattered reports exist concerning the later personality development of the twins in question as a possible result of their differential early training; research published in abstract by Farber and Russell (**3**, **8**) suggests, actually, transitory effects upon learning of rather severe failure, albeit with adult subjects.

An additional illustration of this criticism is supplied by Baker (p. 168): " . . . Sheldon notes some parallels between body type and temperament, which throw new light upon the general nature of temperament and disposition." There is no further discussion of this point, which was selected because it is common to others of the texts. Again (p. 322) Baker implicitly attributes schizophrenia to heredity.

A third common flaw of these volumes is a failure to distinguish between "guidance" or "social re-form" writing and writing which is definitely research based. This is done in these books in two ways: frequently legitimately, with clear indications given that the author is speaking from common sense, or his own beliefs, or his own experience. Such usage is common in Ganz, Wolffheim, Jersild, and Peck. In other cases, research is incompletely presented, then socially valued passages follow, giving the unsophisticated or unwary reader the illusion that he is perusing scientific material. Illustrative of this not uncommon flaw, and by no means the most extreme of the possible illustrations, is Jersild's section on prejudice in Chapter 9. The section is faulty scientifically, but is a strongly written social document with which, ideologically, few readers would disagree. Jersild does not make clear, however, that the evidence in the prejudice area is conflicting, nor that personality-prejudice relations are most tenuously established. He has omitted some relevant recent research. In other words, considerable lecture amplification of the section would be required.

Another flaw, less conspicuous in the "application" volumes than in most of the texts, is the standard treatment of fantasy and daydreaming. In what is commonly considered to be an "extroverted" culture such as ours, these behaviors are commonly condemned at least by implication. Most of the textbook authors have adopted such a point of view; none of them has considered Lewin's (7) original and provocative handling of the subject, where fantasy and level of unreality become important behaviors for research and their necessity for normal development is systematically mapped out. Their economy function as set forth by Freud and implicitly by Sears has been likewise neglected.

Few of such criticisms can be made of the following volumes: Martin and Stendler, Moustakas, Stolz and coworkers, and Zubek and Solberg.

With greater frequency, these criticisms apply to the following volumes (not to a degree implying rejection of the book as a whole, but simply in the sense of an alert lecturer being required for certain sections): Crow and Crow, Cruze, Hurlock, Jersild, and Peck.

Flaws such as summarized above appear most frequently in the books by Baker (in his general, not his practical sections), Davidson and Fay, and Halpern. A reader should be extremely alert in his perusal of these last two volumes.

Because of their intention such criticisms are not particularly relevant to the Ganz and Wolffheim books.

Specific Reviews

Crow and Crow in their *Child Psychology* present what they purport to: a standard outline of material in the area in greatly condensed form. There are reference tables keyed to 14 general topics as covered in 17 standard texts in child psychology and development. The organization is topical. The book is sound enough within the limits the authors have set for themselves, although it suffers from condensation and lack of qualification.

Jersild's *Child Psychology* is by and large an admirable book. One feels, reading it, that Professor Jersild has become sufficiently satiated with the vast body of unrelated material in the area that he has struck out determinedly to look for relating concepts (in his case a version of self-theory). He has not used it consistently and rigorously in all sections of this revision. Since the theory suffers from some inadequacy of definition, it could probably not be so used. But for this field, the reviewer believes that "half a loaf is definitely better than none."

In Jersild's chapters on emotional development, he makes the most concerted attempt to integrate the material by means of his self theory. These chapters are in part unabashedly non-data bound, and are full of provocative clinical hypothesizing. Jersild, one might almost say, has provided the first new paradigm for child rearing (the "development of the self" paradigm) since the Freudian. Approached as a research source rather than as gospel, such a paradigm could be helpful to the field.

Martin and Stendler's *Child Development* is an excellent volume in general. There is striking use of pictorial and verbal illustrations, including delightful cartoons. The last three of four topical sections are superior, the first more standard. Scientific caution is generally exercised; rather fresh sociological and anthropological material is introduced in these latter sections; and essentially sound theoretical treatments of the process of socialization and personality formation are included. The authors espouse no particular theory. The book is well designed for intelligent lay readers and there are sections from which the professional reader can gain both pleasure and intellectual reward.

Peck's *Child Psychology* is on the whole a delightfully written, very simple book with as wide a range of interesting and illustrative anecdotes about children as the reviewer has seen.

It is a practical, guidebook type of volume, presumably for undergraduates. Space precludes any relaying of specific content criticisms, of which there are several. There is, throughout the later sections of the book, a heavy emphasis on the psychologist as the potential solver of all child and child-adult problems. The role of other "remedial" professions is perhaps underplayed. The book is very sound for being as simply and pleasantly written as it is.

Zubek and Solberg's *Human Development* is a scholarly and professional work. Written, as the authors state, with a physiological (but non-theoretical) emphasis, it is nevertheless psychological in a constructive sense. Psychological and social determinants of behavior are handled in a sophisticated manner. The book is readable, well organized, moderate. Research is critically considered and conclusions are tempered and sound. Considerable material not in common currency in the United States is introduced and the chapter on motor development is particularly outstanding. Intellectual development is well handled, with some new material introduced. Classical and recent animal research is adduced in the intellectual development and other chapters. The authors seem most at home as well as most creative in the physiological, sensory, motor, and intellectual areas (through Chapter 10 of the total of 15 chapters). The last five chapters (emotional, social development, etc.) are more traditional in their content and organization. The book is a constructive addition to the literature.

Hurlock's organization of *Developmental Psychology* is basically cross-sectional (each chapter takes up a range of topics for a given age range). This has led to one amusing chapter title sequence: Chapter 13, "Old Age (Sixty Years to Death)" is followed by 14, "Old Age (Continued)." In contrast to Jersild's optimism, Hurlock takes as a theme Hollingworth's quote, "The ultimate goal of development is death" (p. 1).

The last section of Chapter 1, "Happy and Unhappy Ages in the Life Span," is one of the most remarkably pessimistic summaries of life since Dostoevsky. A hereditary emphasis is adopted for personality. But on the other hand, "No longer is it necessary to adopt a 'wait and see' policy in the training and education of children, as it was in the past. Instead, it is now possible to plan ahead in the training of children. This puts education and child training on a firmer foundation than it ever has been before" (p. 18). This quote speaks for itself.

Generally speaking, the later or maturity and aging sections of the book are more appealing than the earlier ones, and throughout the problem of sexual development and behavior is well handled. The material in earlier sections is standard, with some overlapping due to the cross-sectional organization. There is some lack of critical discrimination between studies.

Cruze's *Adolescent Psychology and Development* is standard fare, topically arranged and generally not unsound. The "Sturm und Drang" or "out-of-the-continuum" nature of adolescence is stressed. It is assumed that the reader will have had no previous training in psychology. For instance, the physiological bases of "emotion" section rests entirely upon a detailed treatment of the Cannon theory.

Baker's *Introduction to Exceptional Children* is a revision of his 1944 volume. The book is a point of departure for a course in special and remedial education. The sections on the rapid learning and the mentally gifted supply an emphasis frequently lacking in the special education area. Dr. Baker shows his greatest strength when dealing with practical issues.

Davidson and Fay's *Phantasy in Childhood* contains many reports of children's fantasy, viewed according to Melanie Klein's "breast-guilt-infantile-masochism" version of psychoanalytic theory. The children whose fantasies are related in the book apparently all belong to extremely intellectual, Kleinian oriented families in the British Isles. Instructors in the field of child development or clinical psychology can find interesting illustrative fantasies here, and the book contains certain practical tips for guiding children. There is also some interesting material on the development of fantasy.

Moustakas' *Children in Play Therapy* contains primarily therapy protocol from play sessions with normal or moderately "maladjusted" young children. Moustakas is a nondirectivist, theoretically and methodologically. The book should be read with the Davidson and Fay book reviewed above. The contrasts in content of the reported children's verbalizations are very sharp and delineate, at least by inference, the influence of the reporter or "stimulator" of the child's verbal behavior. The bulk of the protocol material concerns one case, and is excellent illustrative material for teachers or workers in the field. There is little dogma or overgeneralization in the book. It is practical, and the children's protocols sound precisely like children in play sessions. It is one of the few sources of detailed play protocol for normal children.

Halpern has written *A Clinical Approach to Children's Rorschachs.* If students must learn to give Rorschachs to children, this book provides a number of protocols and condensed histories and diagnoses. The book partly reviews itself: "What is offered here is what has 'worked' . . . The question arises whether it might not be advisable to delay this communication until all the suggestions offered are subjected to rigorous experimental investigation. The answer is that publication would never occur . . . methods for validating certain aspects of the Rorschach procedure and the meaning of many of the test factors are yet to be discovered" (p. vii).

The procedure advocated by Halpern for the Rorschach, although it reveals her own wide experience, is sufficiently loose to allow the examiner's theoretical or personal biases infinite leeway in determining the results of the test. The primitive, "direct analogue" Rorschach interpretation method is espoused by the author: "The center figure alone, like all midline figues, is extremely likely to be a reflection of the way the individual sees himself" (p. 46). Response modes revealing anxiety and repression are operationally indistinguishable (pp. 58, 59). Literature on such matters as color shock and examiner influences is ignored, although much of it was in print at the presumed time of writing. Statements about shading responses and anxiety are confusing; schizophrenic children are said to be without good form responses, but so also are *all* children of certain ages.

Stolz and her co-workers (*Father Relations of War-Born Children*) have broken much ground in the almost totally unexplored territory of father-child relations. Any worker in the field should read the monograph for this reason if no other. Briefly, the study uses interviews, observations, and projective techniques with mothers, fathers, and children. There are 19 families in the experimental group, and variable control groups depending on the area of study. Essentially, fathers away from home at the time of and for more than 10 months following the birth of their first child seem to be "problem" fathers and children thus launched into life seem to have more difficult adjustments, at least in their early years, than children born into "complete" families. Some criticisms of design, particularly group matching, could be made, but notwithstanding, the results are convincing. Some of the "clinical" and interview material is vividly illustrative. Interpretation is in general conservative.

Ganz (*The Psychology of Alfred Adler*) and Wolffheim (*Psychology in the Nursery School*) have written two interesting psychoeducational books. These are not primarily intended as a contribution to scientific literature, but represent efforts on the part of the authors to apply principles which for them are consistent with the theoretical formulations of Adler and Freud, respectively, to educational practices. Ganz writes with reference to the public schools of Vienna, the elementary and secondary levels, after the first world war, Wolffheim of nursery schools. The books should be read together.

Their contrasting principles can be roughly summarized thus: Ganz: guide children; don't just let them flower. Not immediate experience but eventual adult adjustment is the important thing. Guidance is done through letting children learn through experience the consequences of their own behavior. Wolffheim: Children must be helped to adjust through finding substitute satisfaction for their wishes. They should not be dominated, even by love. Implicitly, catharsis alone is a helpful mechanism for sound personality development. Formal organization has no place in nursery schools. Teachers must consider and recognize their own motives for working with young children. These motives are likely to be complex.

Additionally, Ganz's book, written in 1935 but only recently translated, is a moving historical document.

Postscript

The preceding pages indicate that, in the reviewer's opinion, there are some bright and some relatively dark spots in the child psychology field as it is represented by these 14 volumes. He believes that a more encouraging picture of the field of child psychology (i.e., one more characterizable as progressive and alive in the scientific and philosophical sense) is presented by the current journals, the program of the 1954 American Psychological Association, and critical and evaluative comments as they have appeared in relevant chapters in the *Annual Review of Psychology*.

REFERENCES

1. ALBERTS, C. A., & EHRENFREUND, D. Transposition in children as a function of age. *J. exp. Psychol.*, 1951, **41**, 30–38.
2. CARMICHAEL, L. (Ed.) *Manual of child psychology*. New York: Wiley, 1954.
3. FARBER, I. E., RUSSELL, W. A., & ANDREAS, B. G. Effect of failure upon performance in verbal and motor learning situations. *Amer. Psychologist*, 1949, **4**, 351. (Abstract)
4. HARLOW, R. G. Masculine inadequacy and compensatory development of physique. *J. Pers.*, 1951, **19**, 312–323.
5. KUENNE, MARGARET R. Experimental investigation of the relation of language transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
6. LAMBERT, W. W., SOLOMON, R. L., & WATSON, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39**, 637–641.
7. LEWIN, K. *Dynamic theory of personality*. New York: McGraw-Hill, 1935.
8. RUSSELL, W. A., & FARBER, I. E. Retention of verbal material as a function of degree of failure experienced in original learning. *Amer. Psychologist*, 1948, **3**, 355. (Abstract)

Maslow, A. H. *Motivation and personality*. New York: Harper, 1954. Pp. xiv+411. $4.50.

Jones, M. R. (Ed.) *Nebraska symposium on motivation, 1954*. Lincoln, Nebr.: Univer. of Nebraska Press, 1954. Pp. x+322. Cloth, $3.50. Paper, $3.00.

To be properly appreciated Maslow's book must be read as a protest, an eloquent and at times almost prophetic protest against the traditional homeostatic conception of drives which has held American motivational theory in a strait jacket ever since our psychological forebears first read Darwin. Maslow rightly sees that even men as apparently different as Freud and Hull were essentially the same in one respect. They both were tremendously influenced in their outlook by the doctrine of evolution, and they both accepted the view that motives were essentially tensional states that the organism tried to eliminate. Maslow objects to such an oversimplified view. In particular, he wants to stress what might be called the "day view" as compared to the "night view" of personality. Thus, he is more interested in need gratification than in need deprivation, in self-actualization than in neuroticism, in healthy rather than sick people, in purposeless behavior rather than instrumental behavior, in art rather than work, in laughter rather than anxiety, etc. Many of the chapters have appeared before in psychological journals and some of them have already influenced current thinking about motivation. As a catalog of deficiencies of traditional motivational theory, they probably have no equal, and they should certainly be read by those graduate students (and perhaps only by them) who feel that they can compass the facts of motivation easily in some simplistic deductive way from a few simple primary drives. It will only anger such students, but perhaps anger is the first step to doing something about the problem.

Like most protests, I feel that it falls short of eloquence on the constructive side. Serious motivational theorists have long known that the problems Maslow describes so eloquently and lists *in extenso* in an appendix as worthy of research ought to be investigated. The problem is not naming them but discovering how they can be attacked both empirically and theoretically. His research contributions, like his clinical study of self-actualizing, extraordinarily healthy people, leave one a little less than satisfied. He cannot, for obvious reasons, name the contemporaries he has considered to be self-actualizing, but he can list some historical figures. Here trouble begins. Someone is sure to point out that nearly all his really "healthy" people are "liberals" or heroes of the liberals, from Lincoln and Debs to Eleanor Roosevelt. Perhaps it is true that only people of certain political convictions can be truly self-actualizing or that self-actualizing people tend to have certain political opinions, but just raising the question makes one wonder whether the judgment of self-actualization depends in any way on one's own values or on the cultural era in which the judgment is made. In short, I think the idea of studying healthy rather than sick people—an idea which Murray first put into practice over twenty years ago—is an excellent one, but it involves some tough theoretical issues that Maslow seems to avoid rather too easily. Those hardheaded scientists who are sure to be annoyed by his eloquence and the cavalier way in which he treats hard theoretical problems would get more out of the book if they treated it as a protest against current theorizing rather than as a closely reasoned contribution to the construction of an alternative theory. As a brilliant protest, it is full of stimulating ideas which, if we can find ways of absorbing them into our science, are bound to make an important contribution to the psychology of motivation.

It is particularly interesting to juxtapose the current *Nebraska Symposium on Motivation* with Maslow's book. To what extent are the points Maslow makes applicable to the six contributions to this Symposium? The comparison demonstrates that some of what he has to say is already dated. Appropriately enough the Symposium opens with a contribution by Farber which represents, more or less, the traditional drive-reduction theory at which Maslow aims most of his artillery, but it ends with a paper by Nissen, who, on the basis of his work with primates, objects just about as strongly as Maslow does to traditional psychological theory about motivation. It is perhaps important to note that both Maslow and Nissen, as well as other critics like Hebb and Harlow, have worked primarily with primates rather than with rats at the animal level. It begins to look as if the theory one accepts depends on the animal one studies. As Maslow says, though, if theorists must pick an animal to think in terms of, why choose the wolf (as Freud does)? Why not the doe or the friendly chimpanzee? In all seriousness, it begins to look as if the rat, which may be an ideal laboratory animal for studying learning, is very inadequate as far as studying motivation is concerned because of its extremely limited repertoire of innate appetites and aversions. And certainly species differences are much greater in motivation than in learning.

But the critics do not take all the honors. A comparison of Nissen's paper with Farber's or with Ritchie's carefully reasoned treatment of the thirst drive in terms of symbolic logic inevitably suggests that while criticism is a part of science, it is not science building itself. One wishes somehow for a middle ground in which the objectors to current motivational theorizing would do some careful system building of their own. In fact, the preference for a tight and orderly system which is obviously incorrect in some details vs. the preference for a flexible approach to concrete facts without too much system strongly reminds me of the distinction Klein makes in his contribution to the Symposium between the "constrictive controls" and the "flexible controls." What science needs is someone who can be both flexible and constrictive. Not too surprisingly I find that Atkinson's extraordinarily complete summary of recent work on the achievement motive seems to avoid overconstrictive "blindness" on the one hand and flexible confusion on the other.

Taken as a whole, the Nebraska Symposium is impressive evidence that the study of motivation is in a vigorous and healthy state. There is not only the important controversy between the traditional drive-reduction theorists and the newer "instinctivists" but there is also a wealth of significant empirical research on different kinds of motives. For example, Festinger reports here some concrete data on the drive to know accurately one's environment, including one's social environment. All that I miss in this exciting research is a concern for measuring individual differences in this particular motive. Atkinson reports extensive research on the achievement motive, carefully integrating it with expectancy theory. Farber is pressing ahead with a questionnaire measure of the anxiety drive, although he is not sure that it is a drive or that he has a measure of it, and neither am I. Klein is working empirically at the interaction point between what Farber would call habit and drive, or what he, himself, would call drive and the method of cognitive control. Nissen has provided us with much valuable concrete observational material on the drives that seem to activate primates. Ritchie is kibitizing on how we should go about our research, and kibitzers have a place in science, irritating though they may be to those who think they are getting along quite well without advice as to how to do research. Furthermore the Symposium permits several of the authors to comment on each other's papers—a most useful idea. In fact I found myself regularly reading the comments before reading the paper itself. For one who grew up in a day when psychology was practically defined as being coterminous with the psychology of learning, I find all this activity gratifying, and it should lead to important improvements in our ideas about human motivation.

David C. McClelland
Wesleyan University

Estes, William K., *et al. Modern learning theory. A critical analysis of five examples.* New York: Appleton-Century-Crofts, 1954. Pp. viii +379. $5.00.

This book is the product of the Dartmouth summer conference on learning theory, held in the summer of 1950 under the auspices of the Social Science Research Council, with funds supplied by the Carnegie Corporation.

Each of the five sections is devoted to one learning theorist (Hull, Tolman, Skinner, Lewin, and Guthrie). The authors have followed a guiding outline for evaluating each theory, including a discussion of the structure of the theory (its delineation of empirical area and its theoretical constructs), methodological characteristics (types of formal model used, techniques of derivation), and, finally, an over-all appraisal of empirical content and adequacy. According to a division of labor among those who attended the conference, individual authors or pairs of authors took responsibility for a given theory, and they have taken responsibility for the sections as they are now written, though of course what is included must reflect to some extent the give and take that went on at Dartmouth. There is no comment on the separate chapters by those who did not author them, so if there were disagreements or minority opinions, these are not reflected in the book.

The chapters move at a high level of sophistication with respect to the logic of theory construction. While the criticisms are often negative in tone, the intent, as indicated in the introduction, is "entirely constructive," that is, the authors look forward to the advance of learning theory beyond the temporary stage that any one contemporary theory must represent. The book should prove very useful to the advanced student. The beginner, who is making his first acquaintance with the theories, might be troubled by the extreme scepticism, and become cynical about the possibility of doing anything of consequence in such a muddled field. He needs additional acquaintance with the ingenuity which these same theorists have displayed in their empirical work, even though their general theories are vulnerable to attack.

Because a critical review of these chapters to be adequate would have to be long, I shall attempt no more than a brief characterization of each. The issues raised in the book will undoubtedly be challenged and discussed separately by those friendly to the various theories.

Hull (by Koch). This 176-page review constitutes nearly half the book. It is based largely on *Principles of Behavior* (1943), although the changes represented in *Essentials of Behavior* (1951) are recognized and discussed, and there is an appended note on the posthumous book, *A Behavior System* (1952). It is unfortunate that the last book could not have been given more adequate treatment, but Koch believes that the 1943 version of Hull's system is likely to remain as the most influential one.

Koch pretty successfully demolishes any claim that Hull succeeded in doing what he set out to do, to develop a system with intervening variables "securely anchored on both sides to observable and measurable conditions or events." He concludes that there is no single case in which intervening variables are actually securely anchored, and that it is impossible to relate any one intervening variable to another "with sufficient determinacy to permit quantitative passage from one to the other" (p. 160). Nevertheless, Hull is credited with doing more than anyone else to put this kind of enterprise on a realistic basis. We do most honor to him if we learn the lessons from his failures, and move on with his vision as to what a science of behavior ought to be.

Tolman (by MacCorquodale and Meehl). The chief difficulty in dealing adequately with Tolman's system is that it has remained programmatic for 20 years, lacking the specificity which makes a more pointed criticism of Hull possible. This lack is compensated for in part by an ingenious effort on the part of the authors to provide a postulate set coherent with Tolman's position, an S_1-R_1-S_2 psychology, in which a stimulus (S_1) eliciting a response (R_1) leads to the expectation of a stimulus (S_2).

Skinner (by Verplanck). Reliance is chiefly on the system as expounded in *The Behavior of Organisms* (1938), and nothing after 1950 is quoted. The purely descriptive outlook proposed by Skinner makes it proper to call his position a theory rather than a system; it is characterized as formal but not formalized and as quantifiable but not quantified. A kind of paradox is detected in respect to prediction:

> . . . it is possible to find no predictions at all of the behavior of rats, or of pigeons, when novel combinations of stimuli are presented to them in the Skinner-box, and many predictions among Skinner's writings with respect to human behavior in a social environment (p. 311).

In general, Skinner avoids generating propositions that may be put to experimental test. Hence some of the criteria of system, applicable to the other writers, are inapplicable to Skinner, at least from his own point of view.

Verplanck appears a little hasty in attributing the success of Skinner and his followers in animal training to "engineering-wise" control, and in considering the successes irrelevant as tests of the adequacy of the theory (p. 272). He similarly believes the experimental productivity to be a tribute to the systematist rather than to his system. One might wish a somewhat further exploration of the possibility that Skinner's position is more closely related to these successes than Verplanck implies.

Lewin (by Estes). The general conclusion is that Lewin failed to construct a field theory, because he was satisfied with a symbolic representation of a situation, and failed to seek precise relationships between variables. Hence there is no adequate field theory to criticize. The ultimate criticism that the theory has not influenced learning experimentation is not quite fair to it. The many experiments on level of aspiration, on the remembering and forgetting of unfinished and finished tasks, on ego involvement and task involvement, on the resolution of conflict, can be classified as learning experiments, and they arise out of the kinds of preoccupations that Lewin's point of view engenders. The possibility must be considered that the area staked out by the typical S-R experiments is too narrow, rather than that Lewin is "too different."

Guthrie (by Mueller and Schoenfeld). Much of the discussion concerns Guthrie's failure to specify clearly the basic terms in his theory, namely, stimulus and response. The content of the system turns out not to be simple, but actually rather formidable; the authors believe that "many reviews in the literature have mistaken incompleteness for simplicity" (p. 368). The failure of Guthrie's account to become more rigorous since he first announced it leads to conjectures as to why it remains prominent. The authors make four points: (*a*) a monistic system such as his is appealing; (*b*) he and his students have been good critics; (*c*) he hit upon some ideas which others have developed, for which he then (deservedly) gets the credit, and (*d*) his theory flourishes in part because reinforcement theory flourishes alongside it, sometimes contrasting with it, sometimes indistinguishable from it.

The critical task which the authors set themselves was admirably done, and we may be thankful for it. The homogeneity in conception of what constitutes a good theory reflects in part the homogeneity of background of those who found it congenial to meet together. They undoubtedly felt heterogeneous enough, yet all of them had been immersed, at one time or another, in the theories of either Hull or Skinner, and in operationism or logical positivism. That this did not produce a tempering of criticism is shown by the severity of the judgments upon the theories of Hull and Skinner, but even so it must have influenced the tasks the conference members set themselves, and the enthusiasm with which they carried out these tasks. Let us grant that a different book would have been written had a different conference been assembled, with theorists initially more sympathetic to other viewpoints. There is nothing to prevent such a new conference, and if it does as well as this one, it will have been of genuine service to contemporary psychology.

ERNEST R. HILGARD
Stanford University

TINBERGEN, NIKO. *The herring gull's world.* New York: Frederick A. Praeger, 1953. Pp. xvi+255. $4.00.

This book summarizes in nontechnical language a large number of observations and experiments by the author and his students on the behavior of herring gulls. Primary emphasis is placed upon social behavior including formation of breeding pairs, establishment and maintenance of a territory, courtship and mating, and rearing of young. Many of the experiments have been reported before and some of them will be familiar to American readers of the international journal *Behaviour*. The theory of behavior which is developed in skeleton form in this volume has been set forth more explicitly in Tinbergen's earlier book, *The Study of Instinct*.

The Herring Gull's World achieves several objectives very effectively. First, it presents a clearly drawn picture of the behavior of gulls in their natural environment. Second, it illustrates authoritatively a method of analyzing behavior which differs in several important ways from techniques used by American psychologists. Third, the book exemplifies an attitude or philosophy of behavior study quite unlike that of experimental psychologists.

Concerning the empirical findings little can be said in a brief review. They are of considerable intrinsic interest and would supply any lecturer in psychology with a wealth of amusing and useful illustrative examples. It seems, for instance, that the story of gulls carrying clams to a great height and then dropping them on rocks which shatter the hard shell and leave the meat exposed is entirely true; and this behavior is not uncommon. However, it is equally true that birds will drop the bivalves on soft mud and repeat this ineffective response a dozen or more times in succession.

The methods described in the book differ from the typical American approach to animal behavior in several ways. In the first place, all observations are made in the field under conditions involving the least possible interference with the animal's normal activities. In the second place, specific reaction patterns such as the brooding of eggs or feeding of young are not studied as isolated bits of behavior, but are analyzed in terms of their relation to all other elements in the bird's reproductive performance and the total behavioral repertoire. Although primary emphasis is placed upon the prolonged and detailed observation of undisturbed gulls, simple controlled experiments are often conducted, particularly in connection with attempts to discover the sensory stimuli involved in various responses. Thus, the sensory basis for retrieving of displaced eggs was determined by studying the reaction of broody pairs to artificial eggs of different sizes, shapes, and colors. The stimuli that evoke the feeding response in which young birds take food from the parent's bill were determined by presenting nestlings with models of adult heads having bills of different shapes and markings. In these experiments, incidentally, it was found possible to create artificial eggs and imitation heads which elicited the normal reaction even more effectively than did the normal stimuli.

The approach to behavior exemplified by Tinbergen's latest book is in general teleological and phylogenetic. It is representative of the philosophy of those students of behavior who call their specialty "ethology," and who concentrate primarily upon the analysis of behavior of wild (undomesticated) animals living in their natural state. The first insistence is upon the development of observational techniques adequate to provide as complete as possible a knowledge of the total behavioral repertoire of the species under examination. Not until such a description is reasonably complete is it considered permissible or expedient to begin a more detailed analysis of separate items in the repertoire such as nest building or food getting.

The ethological approach is teleological in that it involves the assumption that structural and behavioral characters serve some function; and one goal of research is the elucidation of the functional significance of various behavioral responses. This point of view derives quite naturally from a basic concern with the evolution of behavior. Behavior is both an effect and a cause of evolution. Not content with understanding behavior in the herring gull, the ethologist wants to know how it compares with the behavior of other species of gulls, and in particular he is desirous of tracing its probable evolutionary history.

This general interest in the total behavior pattern, its functional significance and phylogenetic background, and particularly the insistence upon protracted field observations seems to generate a special attitude of the scientist toward his animal subjects. Or it may be the other way around—perhaps the attitude precedes and dictates the special interest and method. In any event, many ethologists, and Tinbergen is an excellent example, are emotionally attached to the animals they study. When the scientist is capable of literary expression, his attachment to his subjects lends charm to his writings. This is certainly the case in *The Herring Gull's World*. The author's enthusiasm for behavior study combines with his long-standing affection for sea gulls to produce an eminently readable, entertaining, and informative volume, the attractiveness of which is enhanced by numerous excellent photographs of gull behavior.

FRANK A. BEACH

Yale University

GILLIN, JOHN. (Ed.) *For a science of social man*. New York: Macmillan, 1954. Pp. vii+289. $4.00.

As the title and authorship indicate this is another in the line of recent symposia intended to work toward integration in the social sciences. The plan of this book is unique. Each of three disciplines—psychology, sociology, and anthropology—is represented by two prominent theoreticians. Each of these men discusses some of the theoretical elements which could be held in common by his discipline and one of the sister disciplines (frequently one with which he has also been identified). Thus psychology and sociology are discussed by Newcomb and Parsons, anthropology and psychology by Brewster Smith and Hallowell, and sociology and anthropology by Murdock and Becker. Gillin contributes an introductory and a concluding chapter.

While each chapter is said to be the product of joint authorship, there is relatively little integration, and the authors vary considerably in the manner in which they undertake to fulfill their tasks. For example, Becker and Hallowell devote much of their efforts to wise summaries of the interweaving of the histories of the disciplines they discuss, which they conclude by pointing to the future. Murdock gives us an overview of the strengths and weaknesses of sociology and anthropology; Smith discusses convergences in interests; Newcomb discusses concepts which might provide bridges, while Parsons considers the potential usefulness of various psychological theories to sociology, particularly Parsonian sociology.

In his discussion of "Anthropology and Psychology," Smith stresses the influence of the concept of culture on the general orientation of psychologists, but points out the need for concepts which are more differentiated than "culture" and "personality."

What are the differentiated bridge concepts which psychologists might well use? Newcomb stresses three:

role, consensus, and communication. The concept of role is necessary to sociologists because it points to uniformities at one or another collective level without regard to continuities of specified organisms. "To the psychologist, the role concept is attractive because it points to behavior on the part of individuals," and because the role as perceived by the role taker is an important intervening variable in the prediction of interpersonal behavior. Since role behavior varies with perceived consensus, both sociologists and psychologists are interested in the concept of consensus. The communicative act, Newcomb's third bridge concept, is important as a means of studying interaction processes.

Of equal interest to psychologists is the evaluation of psychological concepts and theories by representatives of other disciplines. Parsons uses at least *six* criteria, with which a useful psychological theory should deal: (1) the relationship of individual behavior to the meanings of objects, especially social objects, (2) an analysis of the elements of structure in the situation of action, (3) the organization of behavior in systems, (4) the structure of situations and personality as organized over time, (5) the distinction between structure and process, (6) the fact that action phenomena are made up of systems within systems.

Using these criteria Parsons sees psychoanalytic theory as coming out ahead of instinct theory, behaviorism, and Gestalt theory in usefulness to sociology.

Hallowell, too, stresses the contribution of psychoanalytic theory but ventures the prediction that the peak of its stimulation to anthropology has passed and that the fruitful influences of the future will come from other lines of current development in psychology, particularly those which help in the understanding of the potentialities for novelty in human life.

There is relatively little integration between chapters, yet, looking at the book as a whole, one is impressed not only by the scholarship of the authors, each of whom seems to have unusual command of his own and neighboring fields, but also with their general good will. In fact the reader is sometimes amused by the general tendency of a writer to be more critical of his own field than of his neighbor's. For example, the anthropologist Murdock, says, "The scientific superiority which sociology enjoys over anthropology is perhaps most obvious in the field of methodology." The sociologist Becker, on the other hand, writing about the limitations of methods says, "Anthropology has thus far been saved from being swamped by procedures akin to those now threatening to overwhelm the sociologist only because of its stress on 'real research' stemming from its long and sound field work tradition."

As Gillin points out, "No one can say with certainty when or if we shall have a comprehensive and integrated science of social man." While this volume does not answer the "when," it enables the viewer to regard the "if" with considerable optimism. While the book does not bring to light a large number of bridge concepts such as those discussed by Newcomb, it does indicate a number of important areas of mutual concern which can be approached in a noncompetitive fashion. This makes for stimulating reading.

A postscript: If you hope to refer to concepts covered by several authors, don't expect any help from the index. There is none.

W. J. McKEACHIE
University of Michigan

Wolberg, Lewis R. *The technique of psychotherapy.* New York: Grune & Stratton, 1954. Pp. xiv+869. $14.75.

Most of the published material on psychotherapy has dealt with theoretical justification, the goals of therapy, generalizations of what should be accomplished and why, but relatively little on specifically how to do it. This is not the case with *The Technique of Psychotherapy.* It provides one of the most remarkably comprehensive discussions of specifically what to do in psychotherapy that is currently available. The range and specificity of the topics covered is truly astounding. At least 38 of the 53 chapters deal with concrete, specific problems in the practice of psychotherapy. The topics covered deal with almost every conceivable question that the beginning therapist might ask an expert. The range includes such topics as estimating prognosis, the therapeutic use of dreams, dealing with inadequate motivation, indications of overly strong countertransference, opening the initial interview, techniques of interpretation, dealing with broken appointments, translating insight into action, handling nonpayment of bills, suggesting termination, inviting the patient back into therapy, dealing with relatives, handling resistance, and terminating unsuccessful therapy. This is only a random sample of the many topics discussed.

This book is remarkable not only in the comprehensiveness of its coverage of specific aspects of therapy but also in the flexibility and freedom of dogmatism of the writer. Wolberg's position appears to be best described as eclectic-psychoanalytic. His actual preference for method, as indicated by the many illustrative sections of recorded interviews and the one complete nine-interview short-term case that is presented, seems to be for a fairly active role for the psychotherapist as contrasted with either the orthodox psychoanalytic point of view or the nondirective point of view. In this sense, his technique is closer to the Adlerian approach than to the Freudian. But Wolberg has something good to find in all varieties of psychotherapy. Insight therapy, catharsis, re-education, environmental manipulation, support, reassurance, hypnosis, dream interpretation, free association, and persuasion all have their place for different kinds of therapy and for different kinds of cases.

It is in this broadness of view, this tolerant eclectic attitude, that lie the weaknesses and strengths of this book. The discussion of different kinds of therapy are descriptive rather than critical or analytical. The terminology of the book is what may now be called the everyday terminology of practical psychiatry which is for the most part a loose terminology and one involving vague and overlapping referents. The theoretical sections are consequently the weakest, and a relatively large section dealing with kinds of therapy for different types of patients verges on typological overgeneralization. Although the book describes in detail what to do, it does not go into why it is being done. It is inevitable that, in a field such as that of psychotherapy where so relatively little is known, there are many unwarranted taboos, superstitions, and stereotypes which are hard to justify either empirically or systematically. It is undoubtedly true that Wolberg falls prey to some of these, but the overwhelming impression of the book is that it is one of good, sound common sense. It does an excellent job of communicating the extensive experience of a skilled and thoughtful psychotherapist. It will be of considerable aid and comfort to the beginning student of psychotherapy and the less experienced therapist, and there is much in it that will be valuable to anyone who practices the art of psychotherapy.

The absence of a systematic personality theory explicitly stated does leave many gaps for the psychotherapist. Questions the therapist will have such as, "Why should he use one technique rather than another?" "When should he use them?" "How can he predict the effect of his behavior?" "What kinds of goals should he set up for himself and the patients?" etc., cannot readily be resolved by reading this book. To do therapy one must have more than the specific, concrete suggestions that Wolberg is able to pass on. Which suggestions to accept and which to reject, when to use one rather than another, and how much of current methods is inefficient are questions which may only be answered now by the psychotherapist by application of a systematic approach to personality which he must obtain elsewhere, and ultimately by research.

Psychologists will be interested in a brief section on "who should do psychotherapy" in which the arguments for and against the practice of psychotherapy by clinical psychologists are stated pretty much in nonjudgmental form. The book is clearly written in a simple and direct style which helps to overcome its considerable length. Four hundred and eighty-one references are given.

To the extent that this book sets out to be a description of the specific, concrete techniques of psychotherapy in current use, it is eminently successful. It will be useful in teaching psychotherapy and as a reference for those who practice psychotherapy. Although many therapists will disagree with one or another of the specific suggestions made by Wolberg for handling particular problems, his suggestions, obviously born out of experience and good common sense, can never be dismissed lightly.

Julian B. Rotter
The Ohio State University

Sarason, Seymour B. *The clinical interaction: with special reference to the Rorschach.* New York: Harper, 1954. Pp. x+425. $5.00.

Clinical psychologists who use the Rorschach technique have been sorely criticized by their more experimentally oriented colleagues for failing to take into account the numerous Rorschach research studies. The use of the Rorschach as a psychodiagnostic instrument still seems to be based upon the original, pre-experimental statements of Rorschach and some of his disciples, with little or no attention paid to those investigations that have attempted to validate such statements. Sarason's effort to handle this problem of integrating research findings with the clinical use of the Rorschach will help answer these criticisms and will bring the Rorschach closer to the main body of contemporary empirical psychology.

The present volume is not, however, a research compendium but is rather a detailed presentation of several carefully selected studies, including some unpublished investigations from the Yale laboratories, as well as the implications of these studies for Rorschach interpretation. Of the multitude of Rorschach studies, the author has selected fewer than one hundred for mention or discussion. One fortunate outcome of this selection is that many studies with equivocal results, inadequate design, etc. are excluded. On the other hand, many readers will search in vain for studies that they regard as especially relevant or significant to the problem under discussion.

According to the author, "This book stems from the belief that practically all clinical problems are concerned with data obtained from an interpersonal interaction and that regardless of how interactions differ certain variables are always operative and must be taken into account" (p. 2). The patient's overt behavior, the instructions, the stimulus complex, the patient's previous experiences in similar situations, and the psychologist's behavior are discussed as such *universal* variables. The author then attempts to analyze Rorschach behavior in terms of each of these variables. For example, each of the usual Rorschach determinants is discussed as part of the stimulus complex; each determinant is first defined, usually following Klopfer's scoring scheme, the classical interpretative statements are then given, followed by a summary of the relevant research. Each such section concludes by suggesting modifications in interpretation and by pointing out unsolved problems that necessitate additional research.

The final section of the volume begins with a rather uneven discussion of psychic determinism, the purposiveness of behavior and the defense mechanisms. The graduate students and professional clinical psychologists for whom this volume seems intended will find much of this section quite elementary. The last chapter presents six complete Rorschach protocols together with Sarason's response-by-response, card-by-card interpretation. This is done in an attempt to spell out the interpretive process and represents a courageous beginning to our understanding of the psychology of the *Rorschacher*.

The author's pro-Rorschach bias occasionally leads him to conclusions that others might find unacceptable. This bias is clearly seen in the treatment of the Rorschach indices of performance under stress (pp. 149–155); Meyer Williams' original investigation is presented here as evidence of the Rorschach's capacity to predict such performance. While the several unsuccessful attempts to cross-validate Williams' findings are also discussed, the reader comes away with the impression that Sarason is rather unimpressed with these failures, and Williams' results are substantially unchallenged. One less favorably disposed toward the Rorschach certainly might come to an entirely different conclusion.

Despite the above criticisms, the present volume represents a significant contribution to the literature of the Rorschach and general clinical psychology. The psychologist who would like to have a more solid empirical basis for his clinical use of the Rorschach ought to regard this book as required reading.

LEONARD D. GOODSTEIN
State University of Iowa

GANNON, TIMOTHY J. *Psychology: the unity of human behavior*. New York: Ginn, 1954. Pp. xii+482. $4.75.

This volume is an introductory textbook of psychology for college students. In addition to a somewhat different organization of the material and a different emphasis accorded to various topics Gannon's book introduces certain theoretical considerations which are not ordinarily found in most textbooks.

The book is divided into four parts: The first part, the shortest, is a general introduction. It treats of the historical development of psychology, its definition, and presents a brief survey of the structure and function of the nervous system. The second, entitled "Reception," deals with sensation and perception. The author, in discussing perception, particularly stresses the neural correlates of perception and the localization of functions in the brain. The third, "Response," contains the discussion of reflexes, drives, and emotions. The fourth and last part, "Integration," is the longest, constituting almost half of the book. The topics included in this part are imagery and dreams, learning, memory and attention, the measurement and nature of intelligence, will and volition, and personality. The book is illustrated throughout with well selected pictures, diagrams, and graphs. The style is simple, the exposition clear, making the text easy reading for the student.

The principal concern of the author is to present all experiences, and all behavior, as integrated manifestations of a single living person. A stress on unity characterizes the book. This emphasis is a consequence of the author's view of psychology, which is reflected in his definition of psychology: "the study of the whole range of human experience and behavior insofar as it manifests the reactions of a single, living person to changes within himself and in his environment" (p. 27). Not only does the author emphasize man's unity, organize the contents of the book around this concept, and endeavor to demonstrate unity in the variety of man's reactions, but he also offers a concept of man to explain this unity of man's behavior. He refers to the Aristotelian philosophy and view of man's nature. Convinced that Cartesian dualism and psychophysical parallelism have greatly harmed psychology, and that what psychology now needs most of all "is some principle of unity" (p. 18), he points to the Aristotelian concept of entelechy and hylomorphic doctrine as a sound philosophical basis for the psychological treatment of man. There will be readers who will object to Gannon's introduction of philosophical considerations, but others will welcome the opportunity of acquainting themselves with the Aristotelian explanation of man's unity.

Later in the book the author examines another theoretical issue of importance for psychology: the problem of determinism in psychology. Gannon believes that the discussion of will and volition belongs to psychology, and that the hostility of modern psychologists toward these topics stems from the influence of mechanism in psychology and the assumption of psychological determinism. "The most damaging effect of this hostility," says the author, "towards volition is the manner in which it has forced psychology to deny facts that it cannot explain in terms of existing viewpoints." "What psychology," he continues, "or any other discipline, must do, if it is to retain its title to respect as a science, is to welcome all the facts that have been observed and then attempt to explain them. To date, the efforts of experimental psychology to explain the will have amounted only to attempts to explain it away" (p. 395). In discussing determinism the author, however, may not have distinguished clearly enough for the student the assumption of determinism as a method of science from determinism as a philosophical doctrine. The entire book is characterized by an effort to present to the student not merely a survey of psychological problems and known facts about man but to provide him at the same time with a concept of man's nature which would serve as a key to the understanding of man as such. It clearly aims to be a guide to the study not of some isolated aspects of man's behavior and experience but of the whole man.

HENRYK MISIAK
Fordham University

BRAATØY, TRYGVE. *Fundamentals of psychoanalytic technique.* New York: J. Wiley, 1954. Pp. xi+404. $6.00.

Braatøy, a slightly off-beat psychoanalyst, writes as a facile essayist, drawing on a vast fund of intriguingly patterned knowledge, often careless with words in his first approximations, but showing profound thoughtfulness and meticulous patience in setting forth his material. The material itself will be of variable interest to most psychologists. For example, many will find it difficult to consider the couch in psychoanalysis as seriously as he does. It should be added that he makes the couch, the position of the analyst, the movement of the analysand's neck muscles, the type of head rest, the advantages of allowing the patient to lie on the floor and kick, the crossing of the legs, and a hundred other details, fascinating material for his unstrained but provocatively serious consideration. He has observed, thought about, formulated and reformulated his ideas against a background of intense interest, broad scholarship, and sound clinical sophistication.

Other themes may, however, spark more response than the analytic couch: The analyst as a focus of interest rather than the patient; the detailed relation between motor, physiological, and verbal behavior; the concept of intricate nonverbal communication; the role of time in personality change; and the distinction between neurotic symptoms and neurotic traits—none of these are new topics, but Braatøy demonstrates unusual clarity in presenting specific implications and concrete meaningfulness.

His approach will have little appeal to those psychologists who are interested only in the most crabbed and minimal conclusions based on the most rigorously denatured evidence. His book is, however, a fascinating development in the gradually emerging *rapprochement* between those analysts who are completely unconscious and those psychologists who permit themselves to think only with the 10 per cent of their iceberg minds that maintains a bobbling existence above sea level.

Braatøy quotes Boring, Eysenck, and Lashley as readily as Ferenczi, Fenichel, and Wilhelm Reich. With equal familiarity he draws on history, philosophy, literature, physiology, and his own intensive experience with patients. He has studied in the Berlin Psychoanalytic Institute, the Salpétrière in Paris, the Menninger Foundation, and chiefly in Norway. He has, moreover, lived with an awareness of customs and people beyond his work with patients. His thinking seems to have grown slowly and carefully; yet he does not try to arrive at an arithmetic sum of his knowledge that equals a given set of conclusions.

While much of the book lacks the authority of firmly established evidence, its purpose is more to consider implications that go beyond the evidence. Not all of the sections are of equal calibre, and the over-all product cannot be offered as an ideal example of "intuitive" writing from a background of wide knowledge. It does often achieve a provocative combination of unusual scholarship and painstaking speculation. Braatøy's early death in 1953 before the publication of the book is regrettable.

ROY M. HAMLIN
Western Psychiatric Institute and Clinic, University of Pittsburgh

GINZBERG, ELI, HERMA, JOHN L., & GINSBURG, SOL W. *Psychiatry and military manpower policy.* New York: King's Crown Press, Columbia Univer., 1953. Pp. xi+66. $2.00.

In 1950, General Dwight D. Eisenhower established the Conservation of Human Resources Project at the Graduate School of Business at Columbia University. The present volume reports one aspect of the continuing study of human resources being made by this group in which the staff of the project brought into focus opinions of a group of distinguished psychiatrists who had attempted to deal with a number of insistent military manpower problems at the request of the project directors.

During World War II, almost a million men were rejected from military service because of psychiatric defects and an almost equal number were separated prior to demobilization because of emotional instability, ineptitude, or undesirable traits of character. It is reported that there is rather general agreement among the psychiatrists queried that the case history of the individual, including his school, work, and social adjustments, is more important than any evidence of tensions shown during a brief medical or psychiatric examination.

A suggestion is made that the psychiatric standards set during World War II were too high. Rejections, it is proposed, might be made on the basis of demonstrated incapacity to adjust to civilian life rather than on any diagnostic label that may be applied to a man.

The question of why men broke during military life is answered in the volume in an interesting way. It shows that psychiatrics have moved far from the simple premise held by some physicians early in the war that each breakdown was a function of a demonstrable personality defect. It is now recognized that stress and "support balance" in the environment are important. Leadership and group morale are also significant in causing or preventing breaks.

Adjustment in military and in civilian life is compared. The replies to this question by the psychiatrists who knew military life intimately were diverse. The authors conclude that one of the important tasks for further research in this area is a better understanding of the factors that contribute to or interfere with a successful civilian as well as military life.

The volume contains many specific suggestions for the saving of manpower. The importance of further research is emphasized.

LEONARD CARMICHAEL
Smithsonian Institution

BRAY, DOUGLAS W. *Issues in the study of talent.* New York: King's Crown Press, Columbia Univer., 1954. Pp. xi+65. $2.00.

The present volume contains a foreword by Eli Ginzberg, director of the Conservation of Human Resources Project at Columbia University.

It is pointed out that the unusually talented person has always attracted the attention of many groups of individuals but that similar study has not been given to the people who range through the whole spectrum of human abilities. In considering the part played by the hereditary potential in general intelligence the author concludes, in briefest terms, that individuals differ in intellectual potential at birth, but that differences in IQ among children from different socioeconomic cultural groups may be due to differences in cultural opportunity. In order to determine the amount of hereditary potential in a population it is necessary, therefore, to correct for differences in cultural opportunity. It is suggested that if cultural opportunities were about equal for all, the percentage of individuals with high IQ's in a favored cultural group might pertain to the population as a whole.

The importance of factor analysis in the study of human abilities is emphasized as is also the significance of studies of motivation and work performance. Much emphasis is given to an analysis of two basic motives, the desire of the individual to do well in school and his desire to reach a particular socioeconomic goal. A design for the study of talent and of superior performance is worked out in the volume.

The present small book will be of interest to those who are concerned with the fundamental question of the causal factors that underlie the individual differences which characterize those who make up our present-day society.

LEONARD CARMICHAEL
Smithsonian Institution

SAPPENFIELD, BERT R. *Personality dynamics.* New York: Knopf, 1954. Pp. xiv+428. $5.50.

Drawing from the best of modern work in the areas of phenomenology, learning theory, field theory, and especially psychoanalysis, the author has done a very skillful job of integrating the whole into a generally readable text on the psychology of adjustment. A strong point in favor of the book is the liberal (albeit sometimes too liberal) quotations from original works and his own very adequate interpretations of them. The text deals with the usual topics of personality dynamics—motivation; developmental aspects of personality; frustration and conflict; aggression and related problems; anxiety, guilt, and inferiority feelings; and repression and other ego-defense mechanisms.

The author has attempted to clarify some of the classical formulations, such as those of id, ego, and superego. Although some may not agree with the exact equating of anxiety with biogenic needs which do not possess acts or objects for their gratification, the attempt to explain the nature of anxiety with these needs as a base is of particular interest. Of greater value is the author's probing consideration of the various explanations of anxiety. He considers the standpoints of orthodox analysis, neoanalysis, maturation, and learning theory, and attempts to integrate them into his own viewpoint. This integration, which provides for an explanation of anxiety as a clearly drawn developmental sequence, is a gratifying contribution to the more serious student who may have been baffled by various attempts to explain this most central concept.

Whereas the style of the text is readable, it is by no means easy. The quotes from original sources, plus the author's own theoretical evaluations, place the book in a dubious position from the standpoint of its use as an introductory offering in courses in adjustment or mental hygiene. Although illustrations are usually adequate, there are times when more liberal usage of them would enhance the discussion. For example, the author states, "The individual develops a self-concept of adequacy or self-confidence when he has had frequent success in regard to positive motives; he develops a self-concept of security when he has had frequent success with respect to negative motives" (p. 36). Although positive and negative motives are discussed, there is room for clarification of this point by concrete illustration. Again, whereas experimental work is often cited, there are times when the author may appear dogmatic, as to wit, his statement (p. 95) that a child who has never had love is less frustrated than the child who has had it and lost it. It would seem that there should be corroborative evidence for such statements.

An impressive asset of the book is the carefully concise summaries at the end of each chapter, and a glossary consistent with textual terminology. There is no doubt that we have here a well thought out volume which is certainly among the best that this reviewer has seen in recent years.

Ralph D. Norman
The University of New Mexico

Ingham, H. V., & Love, L. R. *The process of psychotherapy.* New York: McGraw-Hill, 1954. Pp. ix+270. $5.00.

In the preface the authors state "This book is an attempt to describe the ways in which a psychotherapist works. . . . Anyone engaged in psychotherapy is concerned with both an understanding of people and an appreciation of the ways of dealing with them. Certainly much more consideration has generally been given to a presentation of dynamic theory or to using the advantages of combining both fields than to efforts at describing how psychotherapy is conducted. It is our feeling that there has been a need for a treatise which is centered on the activities involved and is relatively free of other material. The comprehension of psychotherapy and of personality is, of course, not ultimately separable. But it seems to us that it is justified for practical purposes to address ourselves primarily to the procedural aspects. It then follows that we have made no attempt to introduce dynamics of personality except where immediately necessary. . . ."

The authors have been singularly successful in avoiding raising and discussing problems either of "dynamic theory" or the relation of theory to psychotherapeutic practice. It is this success which raises the following question: Can (should?) psychotherapy be taught or learned without some explicit discussion of the principles or dynamics of behavior? For example, how helpful or practical is the following to the student: "As an analyzer one becomes alert to note significant gestures, inflections, and silences. All people give innumerable indications of their feelings, but most of the emotions occur in restrained, abortive fashion in a civilized adult. With practice, though, the therapist can become sensitized to the presence of a great many indicators. When their importance is recognized, then occurrence is surprisingly easy to spot" (p. 73). *How* does the student determine the dynamic significances of silences—on what basis does he deduce the covert from the overt?

It is difficult to evaluate this book because we are not told for whom it is intended. It certainly is not a book for the specialist, and this reviewer is frankly puzzled about the kind and level of student who could benefit from it.

Seymour Sarason
Yale University

BOOKS AND MONOGRAPHS RECEIVED

Ancona, Leonardo. *La psicologia sociale; negli stati uniti d'America.* (Pubblicazioni dell'Universita Cattolica del S. Coure, Vol. XLV) Milano: Societa Editrice "Vita e Pensiero," 1954. Pp. xi+154. 1000 Lire.

Bauman, Mary K. (Ed.) *Adjustment to blindness.* Harrisburg: Division of Documents, Department of Property and Supplies, Commonwealth of Pennsylvania, 1954. Pp. xiii+198.

Bellak, Leopold. *The thematic apperception test and the children's apperception test in clinical use.* New York: Grune & Stratton, 1954. Pp. x+282. $6.75.

Brand, Howard. (Ed.) *The study of personality.* New York: Wiley, 1954. Pp. xvi+581. $6.00.

Coleman, Lester L. *Freedom from fear.* New York: Hawthorne, 1954. Pp. 285. $3.95.

Cooper, L. F., & Erickson, M. H. *Time distortion in hypnosis.* Baltimore: Williams & Wilkins, 1954. Pp. ix+191. $4.00.

Dressel, Paul L. (Ed.) *Evaluation in general eduation.* Dubuque: Wm. C. Brown Co., 1954. Pp. viii+333. $4.00.

Dressel, P. L., & Mayhew, L. B. (Eds.) *Critical thinking in social science.* Dubuque: Wm. C. Brown Co., 1954. Pp. vii+36. $1.25.

Dressel, P. L., & Mayhew, L. B. (Eds.) *General education; explorations in evaluation.* Washington, D. C.: American Council on Education, 1954. Pp. xxiii+302. $3.50.

Dressel, P. L., & Mayhew, L. B. (Eds.) *Handbook for theme analysis.* Dubuque: Wm. C. Brown Co., 1954. Pp. v+78. $1.50.

Dressel, P. L., & Mayhew, L. B. (Eds.) *Science reasoning and understanding.* Dubuque: Wm. C. Brown Co., 1954. Pp. vii+223. $3.50.

Evans, Jean. *Three men; an experiment in the biography of emotion.* New York: Knopf, 1954. Pp. xviii+297. $3.75. (Text Edition, $2.75.)

Gregory, C. C. L., & Kohsen, Anita. *Physical and psychical research; an analysis of belief.* Surrey: The Omega Press, 1954. Pp. ix+213. 15 s. net.

Gross, Feliks. *Foreign policy analysis.* New York: Philosophical Library, 1954. Pp. xxiv+179. $3.75.

Henry, Andrew F., & Short, James F., Jr. *Suicide and homicide.* Glencoe, Ill.: The Free Press, 1954. Pp. 214. $4.00.

Keller, Fred S. *Learning; reinforcement theory.* Garden City: Doubleday, 1954. Pp. ix+37. $.85.

Kardiner, Abram. *Sex and morality.* New York: Bobbs-Merrill, 1954. Pp. 266. $3.00.

Kerman, Edward F. *What is electroshock therapy?* New York: Exposition Press, 1954. Pp. 152. $3.50.

King, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954. Pp. xiv+185. $3.50.

Kugelmass, Newton I. *The management of mental deficiency in children.* New York: Grune & Stratton, 1954. Pp. xii+312. $6.75.

Lewis, Hilda. *Deprived children.* New York: Oxford Univer. Press, 1954. Pp. xvii+163. 9S. 6d. net.

Lindzey, Gardner. (Ed.) *Hand-*

book of social psychology. Cambridge: Addison-Wesley, 1954. Volume I. *Theory and method*. Pp. x+588. Volume II. *Special fields and applications*. Pp. x+601–1226. $15.00 (set), $8.50 (one volume).

LOWENFELD, MARGARET. *The Lowenfeld mosaic test*. London: Newman Neame, 1954. Pp. 349+index (with 144 colored plates in separate container). 50 S. net.

MISIAK, HENRYK, & STAUDT, VIRGINIA M. *Catholics in psychology*. New York: McGraw-Hill, 1954. Pp. xv+309. $5.00.

MURPHY, G., & BACHRACH, A. J. *An outline of abnormal psychology*. (Rev. Ed.). New York: Random House, 1954. Pp. ix+597. $1.45.

OESER, O. A., & EMERY, F. E. *Social structure and personality in a rural community*. (Vol. II of *Studies of social behavior*.) New York: Macmillan, 1954. Pp. xiii+279. $3.75.

OESER, O. A., & HAMMOND, S. B. (Eds.) *Social structure and personality in a city*. (Vol. I of *Studies of social behavior*.) New York: Macmillan, 1954. Pp. xxiii+344. $4.50.

PFUETZE, PAUL E. *The social self*. New York: Bookman Associates, 1954. Pp. 392. $4.50.

PIÉRON, HENRI, *et al*. *L'utilization des aptitudes*. (Livre troisième de *Traite de psychologie appliquée*.). Paris: Presses Universitaires de France, 1954. Pp. ix+416. 1.500 fr.

RÉVÉSZ, G. (Ed.) *Thinking and speaking*. Amsterdam: North-Holland Publishing Co., 1954. Pp. 205. $4.00.

SCHAFER, ROY. *Psychoanalytic interpretation in Rorschach testing; theory and application*. New York: Grune & Stratton, 1954. Pp. xiv+446. $8.75.

SOAL, S. G., & BATEMAN, F. *Modern experiments in telepathy*. New Haven: Yale Univer. Press, 1954. Pp. xv+425. $5.00.

STANTON, ALFRED H., & SCHWARTZ, MORRIS S. *The mental hospital*. New York: Basic Books, 1954. Pp. xx+492. $7.50.

WOLFENSTEIN, MARTHA. *Children's humor; a psychological analysis*. Glencoe, Ill.: The Free Press, 1954. Pp. 224. $3.75.

ZISKIND, EUGENE. *Psychophysiologic medicine*. Philadelphia: Lea & Febiger, 1954. Pp. 370. $7.00.

VOL. 52, NO. 3 MAY, 1955

Psychological Bulletin

PROCESSES AFFECTING SCORES ON "UNDERSTANDING OF OTHERS" AND "ASSUMED SIMILARITY"[1]

LEE J. CRONBACH

College of Education, University of Illinois

How one person judges another is a problem important for its theoretical implications and for its practical significance in group psychology, assessment, teaching, etc. Recent studies of "social perception," as this area may be termed, have been chiefly concerned with differences among perceivers either in terms of their accuracy or in terms of their tendency to view others as similar to themselves.

These studies have usually been built around a particular operation in which a Judge (*J*) "predicts" how another person (*O*) will respond. Often, for example, both persons describe themselves on a personality inventory, and *J* is then asked to fill out the inventory as he thinks *O* did. The extent to which the prediction agrees with *O*'s actual response is taken as a measure of *J*'s accuracy of social perception (or "empathy," "social sensitivity," "diagnostic competence," etc.). Scores obtained in this manner are difficult to interpret, and several investigators have reported distressingly low consistency for them (**10**, **15**, **32**). Although this paper discusses only the perception of Others, many of the findings have relevance also to studies of "insight" where a comparison is made between the subject's self-rating and the rating given him by Others.

This paper seeks to disentangle some of the many effects which contribute to social perception scores, and to identify separately measurable components. This analysis (*a*) shows that investigators run much risk of giving psychological interpretation to mathematical artifacts when they use measures which combine the components, (*b*) directs attention to some especially interesting aspects of social perception left untouched by the usual approach, and (*c*) suggests new ideas regarding the practical use of judgments in clinic, school, and other places.

Our analysis of social perception scores may be instructive regarding research strategy generally. This area of research has developed in an ultra-operationalist manner; of late, workers have seemed content to regard "empathy" as "what empathy tests measure." The principal research activity has been correlating "empathy," so defined, with other variables. We shall show, however, that the operation involves many unsuspected sources of variation, so that scores are impure and results uninterpretable (see also **16**). Studies based on myopic operationism are largely wasted effort when the operation does not correspond to potentially meaningful constructs. Defining a measure operationally is only a preliminary to analytic studies which can refine the measure and bring it closer to the intended construct.

---

[1] Appreciation is expressed to Mary E. Ehart, who assisted in all stages of this paper from initial conception to final interpretation, and to Urie Bronfenbrenner and associates for helpfully providing data and for their courtesy in exchanging ideas throughout our rather similar investigations. This study was conducted under ONR Contract N6ori-07135, of which Fred E. Fiedler is now principal investigator.

Our report on a specialized area of perceptual research shares much of the perspective of Postman's important general review of perception (26). His remarks are peculiarly pertinent to studies of social perception, even though he was thinking more of the "New Look" studies of perception of words and objects:

> At this juncture of debate, we shall do well to pull up short a moment and reconsider the fundamental operations of our perceptual experiments, particularly as they bear on the validity of the theoretical constructs linking perception to motivation and personality. . . . Experiments have shared a common tendency which may be called the projective bias—a selective emphasis on central motivational determinants at the expense of adequate attention to the verbal and motor response dispositions of the subject and the relation of these dispositions to the dimensions of the stimulus. . . . We must then reaffirm the critical importance of a full and precise analysis of the responses as well as the stimuli which furnish the basic data of perceptual experiments (pp. 17–19).

COMPONENTS OF THE ACCURACY SCORE

In the typical experiment we have O's self-description x_{oi} on a set of items, and a set of predictions y_{oij} by J. Error in prediction is represented by the discrepancy between x_{oi} and y_{oij}. An over-all score representing J's ability to perceive others is obtained by averaging his squared errors over all items and all Os. The precise mathematical formulation of this Accuracy score, and the assumptions involved, are presented in the Appendix. It is important to renew here (see 8, p. 457) the warning that any index combining results from

difficulties of interpretation. Whatever factors the items measure, a "global" measure combines with definite weights into a composite. Effects which operate differently on the several factors may be masked. An accuracy score based on heterogeneous items is only an exploratory procedure; where possible it should be replaced or extended by separate analyses of J's ability to predict different qualities of O.

The Appendix shows that the usual Accuracy score is the sum of four components we shall call Elevation (E), Differential Elevation (DE), Stereotype Accuracy (SA), and Differential Accuracy (DA).

Elevation (E). The Elevation component has the form $(\bar{y}_{..j}-\bar{x}_{..})^2$. $\bar{y}_{..j}$ is the average of J's predictions over all items and all Os; it reflects his way of using the response scale. The Elevation component is increased by any difference between J's central tendency of responding and the central tendency of the self-descriptions, for all items and Os combined.

Differential Elevation (DE). Differential Elevation reflects how closely J's average prediction for O corresponds to O's central tendency of response, all items pooled and J's central tendency of response held constant. That is, it reports J's ability to judge deviations of the individuals' elevation from the average. We may write DE in this form:

$$DE_j^2=\sigma_{\bar{y}_{o.j}}{}^2+\sigma_{\bar{x}_{o.}}{}^2-2\sigma_{\bar{y}_{o.j}}\sigma_{\bar{x}_{o.}}r_{\bar{x}_{o.}\bar{y}_{o.j}}. \qquad [1]$$

The variance $\sigma_{\bar{y}_{o.j}}{}^2$ expresses J's report of how much Os will differ in elevation. This assumed dispersion in elevation will later appear as a component of the Assumed Similarity score. $\sigma_{\bar{x}_{o.}}{}^2$ is the true dispersion in elevation. The correlation $r_{\bar{x}_{o.}\bar{y}_{o.j}}$ (to be symbolized DEr) represents J's ability to judge *which* Os rate highest on the elevation scale.

In some tests, central tendency of

response (elevation) reflects insignificant response sets. In other tests, elevation has an important psychological meaning. Thus, if a high score on each item indicates morale, the correlation *DEr* shows how well *J* can judge which *O*s say they have the highest morale.

Stereotype Accuracy (SA). Stereotype Accuracy describes *J*'s ability to predict the norm for *O*s. It might be called "accuracy in predicting the generalized other" (3). This score depends on *J*'s knowledge of the relative frequency or popularity of the possible responses.

We may write:

$$SA^2 = \sigma_{\bar{y}.ij}{}^2 + \sigma_{\bar{x}.i}{}^2 - 2\sigma_{\bar{y}.ij}\sigma_{\bar{x}.i} r_{\bar{y}.ij\bar{x}.i}. \quad [2]$$

Here each variance is computed over items. The variance $\sigma_{\bar{x}.i}{}^2$ is the scatter of the actual means. *SA* represents ability to predict the profile of item means both as to shape and scatter. $r_{\bar{y}.ij\bar{x}.i}$ (Stereotype Correlation, *SAr*) represents accuracy in judging mean profile shape, i.e., the order of item difficulties.

Differential Accuracy (DA). Differential Accuracy reflects ability to predict differences between *O*s on any item. This component is averaged over items. As the Appendix indicates, this component too may be regarded as combining variances with a correlation term. The correlation (*DAr*) measures ability to judge which *O*s have highest scores on the item, when the score is taken as a deviation from *O*s' mean. There is one such correlation for each item.

Removal of the *E* and *SA* components reduces all data to deviations from the group mean, or from the predicted mean. The *DE* component examines ability to recognize individual differences in the first centroid factor underlying the items. The *DA* component pools all remaining factors. There is nothing to prevent estimating further factors among the items by factor analysis, and determining *DA* for each factor separately. This type of cluster score on major factors appears preferable theoretically to the simpler index, $\sum_i DA_i$, and more reliable than the DA_i taken separately (8).

Implications. Seven aspects of *J*'s performance have been separated:

1. Elevation component (*E*)	
2. Assumed dispersion in Elevation	Differential Elevation (*DE*)
3. Elevation correlation (*DEr*)	
4. Predicted variation in item means	Stereotype Accuracy (*SA*)
5. Stereotype correlation (*SAr*)	
9. Assumed dispersion on any item (Elevation held constant)	Differential Accuracy (*DA*)
7. Differential correlation (*DAr*)	

The components are not necessarily uncorrelated. Change in any component alters the Accuracy score.

Surely these aspects of social perception do not all reflect the same trait. A Judge who happens to use the same region of the response scale as other persons (Elevation is small) need not have superior insight. Judging which items have the highest mean seems to require acquaintance with the norms of the group; but a person might possess such knowledge and yet be unable to differentiate accurately between individuals (**16**).

At best, failure to separate these components makes interpretation ambiguous. Chowdhry and Newcomb (**5**) requested group members to predict what percentage of their group would agree with each of many attitude statements. Ability to make this prediction was judged by a difference score, and this score was found to correlate significantly with leadership status. This score, however, combines Elevation and Stereotype Accuracy; until the components are separately measured we cannot rule out the possibility that leaders simply used the correct range of the scale more often than nonleaders. This, in turn, might reflect willingness (or unwillingness) to use extreme percentages rather than any other subtle perceptiveness of specific attitudes. That such effects do occur is shown by Lorge and Diamond, who required judges to estimate what proportion of *O*s would pass ability test items. They found that poor judges were greatly helped simply by being told the difficulty of a few specimen items. "Apparently the difference between 'poorest,' 'mediocre,' and 'best' judges is that the 'best' judges have some experiential referent for the per cent of the population that can pass an item. Giving such referents to the 'poorest' and 'mediocre' judges . . . leads to a significant reorientation of such judgments" (**19**, p. 33). When judges responded only to the items, the best judges had a mean *SAr* of .73 and the poorest, .56. After information on averages for items was given, the same judges had mean correlations of .77 and .73. The difficulty encountered in interpreting the Chowdhry-Newcomb study does not arise in Talland's study of the same problem (**33**) where subjects were asked to predict what ranks will be assigned to certain stimuli. In ranking, elevation and dispersion are the same for everyone, and therefore the scores depend only on *SAr*.

Failure to identify the components of the Accuracy score can lead to artifactual correlations. Only a few of the many examples in the literature need be cited. Norman and Ainsworth (**24**) report a large number of correlations between Accuracy ("empathy") and Assumed Similarity ("projection"). Since the Accuracy score contains Assumed Similarity components, the two scores would necessarily overlap even if both sets of responses are determined strictly by chance. The correlations have no psychological meaning. Dymond (**11**) reported that persons with high Accuracy are also most easily judged. But a person who uses the scale in a typical manner will have a small Elevation component, hence better Accuracy; and other persons will have smaller Elevation errors in judging him, simply because of this typicality. This would happen even if the Judge predicted his responses without ever meeting him! Perhaps social psychologists should take what comfort they can from Bertrand Russell's remark that physicists "have not yet reached the point where they can distinguish between facts about relativity and mathematical operations which may have nothing to do therewith."

COMPONENTS OF THE ASSUMED SIMILARITY SCORE

The *J* is said to "assume similarity" between himself and *O* if *J*'s prediction for *O* differs little from *J*'s self-description. One might study assumed similarity with respect to each *O* separately, but we shall give attention to the tendency of *J* to assume similarity over *O*s in general. The formula for this *AS* score is given in the Appendix. *AS* is sometimes interpreted as "projection" or "identification" (**29**). As indicated in Equation 5*a* of the Appendix, the *AS* score divides into the components Assumed Elevation (*AE*), Assumed Self-Typicality (*AST*), and two Assumed Dispersions (*ADE*, *ADI*).

Assumed Similarity in Elevation (AE). The first component, Assumed Similarity in Elevation, takes the form $(\bar{y}_{..j}-\bar{x}_{.j})^2$. It measures *J*'s tendency to assume that *O*s have the same average response as he does. This component would be important if items are polarized so that a high score on each represents good adjustment or some other interpretable quality; the score then shows whether *J* regards the average *O* as similar to himself in this central dimension.

Assumed Dispersions (ADE, ADI). A second component is $\sigma_{\bar{y}_{o.j}}{}^2$, the Assumed Dispersion in Elevation (*ADE*). Another is the Assumed Dispersion on specific items after differences in Elevation are removed (*ADI*). *ADI* closely resembles Gage's concept of "rigidity" or "adherence to stereotype" in prediction (**14**, p. 16; **15**). These dispersions have already been encountered as components of *ACC* (see Equation 1 above and Equation 4*a* in Appendix). We shall refer to them as Assumed Dispersion in Elevation (*ADE*) and Assumed Dispersion Within Items (*ADI*), respectively.

Assumed Self-Typicality (AST). The remaining component measures the discrepancy between *J*'s perception of the average *O* and his self-description. This component tells whether *J* regards his own profile as typical in shape. Or, we might say, this component shows the similarity of *J*'s self-perception to his implicit stereotype of *O*s (Elevation held constant). We follow Gage (**14**, p. 17) in calling this Assumed Self-Typicality (*AST*).

Of the four *AS* components, only *AST* divides into separate variance and correlational terms, as shown in Equation 6*a* of the Appendix. The correlation represents the similarity between *J*'s self-description and the average profile, ignoring differences in elevation and scatter. We call it the Self-Typicality Correlation (*STr*).

IMPROVEMENT OF PREDICTIONS

Insofar as our mathematical model is an acceptable approximation to real problems, we can reason mathematically to determine how judgments of *O*s may be improved. The conditions which make errors of prediction as small as possible are stated fully in the Appendix. The most significant principle takes this form: Accuracy is improved as σ_y approaches $r_{xy}\sigma_x$. That is to say, the variation in predictions should never exceed the variation in true responses, and should ordinarily be much smaller.

This principle indicates that there is an optimal degree of differentiation in making judgments. If *J* can make accurate judgments as to the relative location of *O*s on a continuum, then he is wise to make σ_y as large as σ_x—never larger. But if he is forced to base his judgment on inadequate cues or if the available personality theory and situational knowledge do not permit trustworthy inference, then he should treat people as if they were very nearly alike. The person who attempts to differentiate individuals on inadequate data introduces error *even when the inferences have validity greater than chance.*

The variation of *J*'s predictions indicates how much he differentiates. For example, a teacher estimating

IQ's in a class might spread them from 90 to 110 or from 70 to 130. We would expect the judge who perceives greater differences to apply more sharply differentiated treatments to the various persons. A person who knows that the expected σ for IQ's is 16 might try to predict so that his estimates would have this σ; but unless he is a perfect judge, this is unwise. He will have smaller errors if his predicted σ is less than 16—how much less depending on the correlational accuracy of his predictions.

If two diagnosticians can each judge some trait with correlational validity .40, the one who differentiates strongly (i.e., makes extreme statements) will make far more serious absolute errors than the one who differentiates moderately. Indeed, the person who makes extreme differentiations based on a validity of .40 may have larger errors than a judge who has zero correlational validity but gives the same estimate for everyone. This contradicts the view that judgment is always improved by taking into account additional valid information.

Implications regarding clinical judgment. Clinical judgments are frequently regarded as undependable, because of research tending to show that clinicians weight predictor variables inappropriately. Thus Sarbin showed that counselors given a great amount of information predicted grade averages no more accurately than did a regression formula, one reason being that the counselors gave the *ACE* test excessive weight (**27**). A similar problem of weighting is involved in attaining the ideal dispersion of estimates. The regression equation for two uncorrelated predictors in standard form is:

$$x_{1.23}=r_{12}x_2+r_{13}x_3+w_e x_e. \qquad [3]$$

The w_e in the last term is a weight for error, and x_e is an individual's estimated error score. We do not usually write this term because x_e is zero and the term vanishes. Since $\sigma_{1.23}=\sqrt{1-w_e^2}=R\sigma_1$, the statistical prediction formula gives the optimum degree of differentiation, using the proper w_e. But the clinician who combines variables 2 and 3 with the best relative weights may still obtain an inaccurate estimate if he employs the wrong w_e, and so makes $\sigma_{1.23}$ too large.

There is evidence, both from Sarbin's study and from similar recent work by R. S. Melton (**20, 21**) that counselors do overdifferentiate. In Sarbin's study, σ_x for grade-point averages was .88 (for both sexes, computed by this writer from Sarbin's report). The statistical predictions had a σ of .47; the clinical predictions, .57. In unpublished data supplied by Melton, σ_x was .59, but σ_y (for predictions) was .60. The optimum, $r\sigma_x$, would presumably have been near .35.

Sarbin was puzzled by the finding that $\sigma_y<\sigma_x$, and he discusses this as a "disadvantage" (**28**, p. 599). His remarks resemble those of others on the so-called "central tendency of judgment," which has hitherto been regarded as a source of inaccuracy in social perception (**1**, p. 521). But estimates from a regression formula have a lower variance than observed results just because weight must be assigned to error. Sarbin's clinicians made some, but insufficient, allowance for their error. Melton's clinicians made no allowance for error. Evidently the fault of the clinician is too little "central tendency of judgment."

Melton's study combines all sources of error into a single measure of absolute error. He reports that counselors make greater error than a statistical predictor, and that this remains true even when the clinicians are given an actuarial table as a guide. Decomposing his accuracy score into components yields further knowledge about the predictive process. From the data he has supplied us, we learn that the group given the actuarial table makes no error in estimating mean GPA, while the control group overestimates by $.65\sigma$. Both groups differentiate far more than the optimum (and this effect appears to be intensified with the actuarial table). From the data at hand, we do not know whether correlation accuracy (*DAr*) is higher or lower for Melton's counselors than for a statistical prediction, nor how the actuarial table altered this aspect of their judgment.

Systematic errors such as overoptimism and overdifferentiation may be corrected fairly easily. It is important for studies of clinical judgment to measure these errors as separate components, and for clinicians to train themselves to avoid these errors.

Implications for teaching. Recognizing an optimum degree of differentiation makes it necessary to reexamine and qualify statements commonly made in training teachers, to the effect that every pupil has his own pattern and the teacher must fit methods to that pattern, not treat the pupil in terms of the statistical average. The writer has himself expressed such views, but it now appears that the teacher who is poorly informed regarding the unique patterns of his pupils should probably treat them by a standard pattern of instruction, carefully fitted to the typical pupil. Modifying plans drastically on the basis of limited diagnostic information may do harm. Differentiation is harmful if the extent of adaptation or differentiation exceeds the amount justified by the accuracy of social perception.

Teachers may properly modify treatments considerably when they are well able to judge individual differences. Differences in arithmetic achievement they might judge quite accurately; if so, they could profitably provide quite different assignments for different individuals. But if it is hard to judge creative potential in art, say, or the ultimate vocational goal of a ninth-grader, then it is a great mistake to differentiate treatment to fit perceived differences.

ILLUSTRATIVE ANALYSIS OF CORNELL DATA

To illustrate our system of analysis, we use data kindly provided by Bronfenbrenner and Dempsey. The data were gathered at Cornell University primarily for pilot analyses such as ours. We shall deal with eight subjects and eight items. The eight subjects were candidates for employment as interviewers. Each person interviewed each of the seven others. In each interview, each man was to obtain information about his partner. Following the interview, each person stated his own reaction to 19 items (of which we use only eight) and predicted what his partner would say. One item is: To what extent did you feel at ease during the interview? ____*a.* very much ____*b.* a good bit ____*c.* only slightly ____*d.* not at all. The respective responses are scored 1-2-3-4.

Completion of the design provides seven self-descriptions and seven predictions by each man (also seven *for* each man). We have taken two simplifying steps which might be illegitimate for purposes other than demonstration. We use the average of *O*'s responses over all seven interviews as his true response, x_{oi}, discarding information on *O*'s variation from interview to interview. Secondly, we treat *J*'s self-description as a "perfectly accurate prediction of

TABLE 1

ACCURACY SCORES OF EIGHT JUDGES DIVIDED INTO COMPONENTS

Judge	ACC_j^2	Elevation (*E*)	Differential Elevation (*DE*)	Stereotype Component (*SA*)	Differential Component (*DA*)	*DE* Contains		*SA* Contains		*DA* Contains	
						$\sigma_{\bar{y}_o.j}$	$r_{\bar{x}_o.\bar{y}_o.j}$ (*DEr*)	$\sigma_{\bar{y}.ij}$	$r_{\bar{x}.i\bar{y}.j}$ (*SAr*)	$\overline{\sigma_{y'}^2}$*	$\overline{r_{x'y'}}$* (*DAr*)
1	2.32	.00	.69	.24	1.40	.24	.18	.47	.93	.12	.28
2	5.32	.05	1.71	1.34	2.24	.38	−.14	.17	.37	.17	.12
3	4.60	.10	.23	.50	3.77	.15	.63	.47	.85	.39	.08
4	5.33	.26	1.35	.91	2.81	.35	.00	.21	.67	.25	.23
5	4.37	.98	.79	.28	2.33	.27	.18	.45	.91	.16	−.04
6	4.06	1.22	.76	.65	1.45	.18	−.20	.54	.86	.11	.26
7	3.75	.20	.73	1.27	1.56	.14	−.44	.15	.44	.17	.38
8	3.78	.02	.41	1.28	2.06	.29	.62	.76	.92	.27	.35
Mean	4.19	.35	.83	.81	2.20	.25†	.10	.40†	.74	.20	.21
Variance	.83	.20	.20	.18	.56	.01	.13	.04	.04	.01	.02

* Averaged over items.

† The values of $\bar{\sigma}_y$, .25 and .40, may be compared to these respective values of σ_x, the true variation: .22, .44.

himself." By this device, we deal at all times with eight *J*s and eight *O*s, and the criterion is made the same for every person.

Accuracy Scores for Eight Persons

Table 1 presents the *ACC* score for each person, and his score on each component. These and subsequent results are illustrative, and not a proper basis for generalization.

Relation of differentiation to accuracy. As expected, any component decreases as the predicted standard deviation (σ_y) approaches the product of the related correlation with the actual standard deviation (σ_x). Consider, for example, the results on *SA* and its constituents. Person 8 is an excellent Judge, according to his *SAr* of .92. But he expects too much variation in the item means (.76 compared to an actual σ of .44). As a consequence No. 8 has a poor *SA* score despite his excellent ability to discriminate between items. The best *SA* scores are earned by No. 1 and No. 5, who have high correlations and who predicted variance close to the actual variance. Compare also *DE* and *DEr* of No. 3 and No. 8. These persons have the same *DEr*, but No. 3 underestimates the variation in elevation, and No. 8 overestimates it. As expected, No. 3 earns the better *DE* score.

The judges consistently overdifferentiate. The optimal σ_y would be about .025 in the *DE* column, and .30 in *SA;* but the actual mean values are .25 and .40. Since $\sigma_x'^2$ =.12, it is clear that $\sigma_y'^2$ is also generally larger than the optimum.

Reliabilities and intercorrelations. Internal consistency was studied for the various components, but the results based on eight cases need not be reported. One finding, however, is notable. Differential Accuracy was strikingly consistent over items: a coefficient of .73 was obtained by analysis of variance. That is to say, some predictors were consistently good over all items, others consistently poor. But when we examine the components of *DA*, we find that Assumed Dispersion Within Items is consistent over items (.79), and *DAr*, the measure of accuracy in differentiating, is not (.18). In this sample, Differential Accuracy shows reliability only because some persons have consistent sets to differentiate. Stone and Leavitt (32) likewise find very low consistency (−.07 to .30) of accuracy scores in predicting different children, but a median consistency of .63 between two predictions for

TABLE 2

ASSUMED SIMILARITY SCORES DIVIDED INTO COMPONENTS

Judge	AS_i^2	Assumed Elevation (AE)	Assumed Dispersion in Elevation (ADE)	Assumed Self-Typicality (AST)*	Self-Typicality Correlation (STr)	Assumed Dispersion Within Items (ADI)
1	1.89	.14	.46	.34	.94	.97
2	3.78	.35	1.15	.95	.60	1.36
3	4.63	.06	.18	1.30	.90	3.09
4	3.45	.16	.98	.30	.61	2.01
5	2.21	.01	.59	.37	.89	1.25
6	1.54	.12	.26	.28	.94	.87
7	2.41	.00	.15	.88	.72	1.37
8	4.15	.12	.66	1.16	.89	2.20
Mean	3.01	.12	.55	.70	.81	1.64
Variance	1.14	.23	.12	.16	.02	.49

* A composite of STr and $\sigma_{\bar{y}.ij}$ of Table 1. See Equation 6a of Appendix.

the same child. They trace the latter consistency to consistent favorable sets toward a given child, and to assumed similarity. All results to date lead us to doubt whether accuracy in differentiating personalities of others can be reliably measured. Where reliable variance is found, it seems to result from some constant mental set.

In Table 1 we note that No. 1 is consistently superior on various components of Accuracy and No. 4 is consistently inferior. But No. 7, the best predictor as judged by *DAr*, is the poorest on *DEr* and next to poorest on *SAr*. With only eight cases, meaningful correlations cannot be obtained.

Future studies of predictive accuracy should measure the components separately, preferably using two independent sets of items and *O*s. Such measurement will permit accurate determination of reliabilities of components, of the relation between the components, and of their relation, if any, to external criteria. Ideally, items would be organized into clusters to permit study of predictions on separate traits. Only after such research can we decide how many components within the over-all Accuracy score presently used are important, and which unwanted components must be suppressed by appropriate design of tests and scoring keys.

Assumed Similarity Scores for Eight Persons

In Table 2, the Assumed Similarity scores are divided into components. The relatively large variance of *ADI* indicates that it has great influence on individual differences in over-all *AS*.

AE correlates .81 with *ADE*, and *AST* correlates .97 with *ADI*. In these data, the tendency to differentiate among *O*s is accompanied by a tendency to differentiate the average *O* from oneself. This result is partly an artifact, resulting from using each person's self-description as one of his "predictions." Even allowing for this, our correlations suggest separating only two components of AE: Assumed Similarity in Elevation (*AE*+*ADE*) and Assumed Similarity in Pattern (*AST*+*ADI*). The correlation between these is only .21. Further evidence is required, however, to establish definitely how to divide Assumed Similarity. An earlier study (**9**) suggests strongly that Assumed Similarity is a general mental set, almost independent of the psychological content of the items. A global index may therefore be satisfactory for this score.

THE JUDGE'S "IMPLICIT PERSONALITY THEORY"

We turn now to an aspect of social perception data which may prove to be particularly significant. When a Judge describes or makes predictions for a large number of *O*s, these predictions define a distribution of points in the variate space. This distribution may be regarded as a description of the generalized *O*, representing *J*'s view of both central tendency and individual differences. The *J*'s generalized perception may be an important indicator of his expectations regarding *O*s. We shall discuss the general significance of this perceptual system before tracing its effect on social perception scores.

The *J*'s distribution is to be examined in terms of the means, variances, and covariances of the predictions. The mean may be regarded as *J*'s stereotype; if the mean *O* in his descriptions is "hostile," for example, this may be highly significant. The variance or assumed dispersion on a variate indicates *J*'s tendency to differentiate along that dimension. The covariance is interpreted as indicating the relation *J* expects to find among variates. A given *J* may customarily report the same persons as high on both "quietness" and "shyness," for instance; or on both "ambition" and "selfishness." These aspects of the distribution reveal *J*'s view of *O*s and the connotation of personality traits for him. We suggest that these means, variances, and covariances describe *J*'s implicit theory of personality.

The expectations *J* has of *O*s constitute his view of personality, and one may hypothesize that they direct his responses to *O*s. G. A. Kelly (**18**) argues that each person forms "personal constructs" by means of which he differentiates situations confronting him, including other persons. The "personal constructs" would be, in our model, the dimensions along which *J* differentiates strongly. As Steiner (**31**, p. 349) notes, it is particularly important to investigate the covariation of his constructs. Osgood (**25**) suggests studying the semantic equivalence of stimuli by testing whether they are used similarly. Our method is quite like his, determining as it does what traits *J* uses to describe the same persons. While Steiner asks a person directly what traits he expects to be associated, we suggest looking at the covariation found among the predictions y_{oi}. Such implicit relations are not subject to deliberate distortion and can reveal associations and norms of which *J* himself is unaware.

An illustrative case. This concept can be illustrated by using a small portion of the Bronfenbrenner-Dempsey data. The *J* predicted responses of eight persons (including himself) on these questions:

1. In general, how openly did you express your feelings and emotions during the interview?
2. How much interest did you feel in the other man as a person?
3. How much were you aware of how he was feeling?
4. How much opportunity did you give him to interview you?
5. How much important information were you able to get about him?
6. To what extent did you feel at ease during the interview?
7. To what extent did you succeed in establishing a good interviewing relationship?
8. To what extent did you feel like the person being interviewed rather than the person doing the interviewing?

The matrix of covariances for Judge 3, a poor predictor, was fac-

TABLE 3

SUBSTANTIAL FACTORS IN JUDGE 3'S COVARIANCE MATRIX*

Item	Factor I "Pressure"	Factor II "Exchange of Information"	Factor III "Rapport"	h^2	Variance $\sigma_{y_{oi3}}^2$	Mean Prediction $\bar{y}_{.i3}$
1	*2.55*	−.68	−.62	7.35	7.58	2.39
2	1.32	−.94	−.16	2.65	3.77	1.80
3	−.76	.16	*.82*	1.28	1.81	1.70
4	.46	*1.35*	−.07	2.04	2.34	1.63
5	.68	*1.17*	.16	1.86	2.16	1.59
6	*−1.92*	−.76	*1.28*	5.90	6.42	1.79
7	−.88	−.30	.63	1.26	2.14	1.43
8	*2.50*	−.08	.40	6.43	6.43	2.93
Per cent of variance	62	16	10	88		
Cumulative per cent	62	70	88			

* Largest loadings in each column in italics.

tored by a pivotal method intended to yield interpretable factors. Table 3 shows the loadings on three factors, and also item means and variances. The means for Judge 3 show no striking features, especially when considered in relation to the true means presented in Table 4. The variances indicate that No. 3 regards others as fairly uniform in their awareness of him (item 3), and as varying especially in openness, ease, and feeling of dominance (items 1, 6, 8). No confidence can be placed in factors based on eight measures, but we would otherwise interpret Factor I as representing a feeling of being under pressure. It is notable that No. 3 regards those persons who are most open (item 1) as being least at ease (item 6). Factor II shows a link between items 4 and 5, getting and giving information. Factor III is indistinct. It is notable that items 6 and 7 are correlated; a "good interviewing relation" is perceived by No. 3 as one where the *interviewer* is at ease! Such a finding regarding No. 3's perception, if better substantiated, might have much diagnostic importance.

Relevant prior studies. The literature contains many studies of correlation between ratings which bear on the perceiver's frame of reference. Reports of halo effect suggest the existence of a strong general good-bad factor. These studies have not examined raters separately. Newcomb (**23**) showed that there were substantial correlations among ratings ($\bar{r} = .49$) even where direct behavioral observations on the same qualities showed a mean correlation of only .14. "The close relation . . . ," says Newcomb, "may be presumed to spring from logical presuppositions in the minds of the raters" (p. 288). Steiner (**31**) found evidence that ethnocentric individuals see others in black and white terms, the "good," "strong" traits going together. In our language, their covariance matrix is loaded with one factor, while nonauthoritarians use many factors and do not emphasize the general evaluative dimension. Soskin (**30**) showed that halo effect and the stereotype profile varied as a function of the data given the assessor and that

TABLE 4

FACTORS IN THE CRITERION COVARIANCE MATRIX DETERMINED BY PIVOTAL METHOD*

Item	Factor I "Openness"	Factor II "Receptiveness"	Factor III "Passivity"	h^2	Variance $\sigma_{x_{oi}}^{2}$	Mean $\bar{x}_{.i}$
1	*2.06*	−.40	−.30	4.49	4.57	2.05
2	.14	*.94*	.01	.88	1.03	1.93
3	.09	*.96*	−.11	.94	1.16	2.04
4	*.86*	.40	.30	1.00	1.07	1.80
5	.22	.55	−.44	.54	.67	2.05
6	.47	.60	*1.16*	1.93	1.42	1.57
7	−.10	*.86*	.09	.76	.97	1.64
8	−.05	−.27	*.85*	.80	2.45	3.09
Per cent of variance	40	27	18	85		
Cumulative per cent	40	67	85			

* Largest loadings in each column in italics.

ratings of peers are distributed differently from ratings by professional assessors.

A striking recent study by Jones (**17**) compares the ratings given by authoritarian (A) and nonauthoritarian (NA) groups to Others regarding whom carefully controlled information had been given. He finds many types of differences including a tendency of the NA's to respond to personal qualities of the Other, whereas A's seem to differentiate less among leaders. The data are analyzed to show, in effect, which traits perceived in the Other vary with the Other's *democracy*. The groups agree in associating democracy with *sensitive to Other*, *generous*, *adaptable*, *warm*. The A's associate *democrat* also with *unambitious*, *poor officer*, *undependable*, *hard to figure out*, *acts without thinking*, *rebellious;* while the NA's associate these qualities with *autocrat*. An attempt was made also to find correlates of *forceful*. The two groups showed little difference in covariances, associating *forceful* with *natural leader*, *ambitious*, *uses his head*, etc. At least two other studies show differences in the perceptual reference frames of groups. Wickman's well-known study (**34**) showed that teachers expected different traits to correlate with mental health than did mental hygienists. Moore (**22**) performed a factor analysis of ratings given noncommissioned officers by their subordinates, and also of ratings given by their superiors. The factor patterns differed. For instance, superiors coupled *leadership* with *eagerness* and *responsibility*, but the subordinates linked *leadership* with *intelligence* and *skill*.

None of these studies of groups examines the perceptual space by which an individual describes personality, but the evidence supports the belief that important individual differences exist. Our proposed analysis of the covariance matrix can be applied to the matrix based on mean ratings given by a group of raters. This should give more complete information regarding group differences in implicit meanings than the methods used in the studies cited. The emphasis in recent studies has been to consider correlations between items as a meaningful phenomenon, related to psychologically interesting qualities of the rater. This may be contrasted to the view in the earliest studies such as Newcomb's, where such correspondences were regarded solely as an annoying interference or so-called "logical error" in rating. It also contrasts with many recent studies which concentrate on the interaction between Perceiver and Other, failing to inquire about elements associated with the Perceiver alone.

The analysis of the implicit meanings of various dimensions for the Perceiver may be used in several ways. If, in teacher training, one aim is to modify the way in which teachers interpret behavior, these changes in viewpoint should be reflected in changes of the perceptual distribution. For example, if teachers naively regard *quietness* as associated with *adjustment*, yet experts regard quietness as unrelated or even negatively related to adjustment in children, then the aim of training is to reduce or reverse the correlation found in the teachers' responses. Because our technique examines implicit interpretations, it should be especially useful for evaluation. Another application of the method is in industrial rating. The rater's distribution, when rating many applicants on many traits, indicates what differences he pays attention to and how he interprets the traits he is supposed to rate. In analyzing one rater in this way, we find that he regards *creative* and *inquiring* as nearly independent of *intelligent*, and their independent contribution is negatively (!) correlated with his final recommendation as to hiring. So far as we can determine, previous studies have pooled all raters before studying trait intercorrelations;[2] the study of idiosyncratic rating patterns should lead to important suggestions for training raters or improving scales. Finally, in view of our interpretation of the perceptual distribution as an implicit personality theory, special interest would attach to studies of ratings given by clinical psychologists or psychiatrists of different schools, who might be presumed to hold different theories.

Effect on accuracy scores. The *J*'s distribution of *O*s has been interpreted here as a standing system of meanings which delimits the space within which he locates *O*s. It is obvious that any such delimitation would affect social perception scores. Discrepancies between mean and actual mean lower Stereotype Accuracy, and Accuracy declines if perceived variance (*ADE*, *ADI*) departs from an optimal value. The correlational effects are a bit less easy to perceive.

Correlations describe the shape of the distribution of *O*s. If traits 1 and 2 are uncorrelated, then x_{o1}, x_{o2} will have a roughly circular joint distribution. If a *J* regards 1 and 2 as correlated, his perceived distribution of y_{o1}, y_{o2} will be elliptical. Perceived variance along the dimension $1+2$ will become greater than in the true responses, and Accuracy will suffer. We can view the example in another way. Suppose the Judge predicts variate 1 perfectly but believes that variates 1 and 2 correlate 1.00—then he must have substantial error in predicting variate 2. He can predict 2 accurately only if he perceives the covariance of 1 with 2 accurately.

Data reported by Crow (**10**, p. 86) show this phenomenon clearly. He asked *J*s to predict what would be the first word missed by a patient on a vocabulary test and what would be the highest level attained (called tasks D1 and D2). The correlation of *J*s' Accuracy on D1 with Accuracy on D2 was positive and significant for five of ten patients, but negative and significant on two patients. Judges tended to *expect* a correlation between D1 and D2, and they were accurate on those patients where the two scores were similar. Where the scores were dissimilar, *J*s could not be accurate on both predictions. There was a rank correlation of .97 (over *O*s) between Accuracy, and consistency of *O*'s performance.

---

[2] Especially clear evidence on this point is provided in a recent dissertation by Walker. (Walker, W. B. An investigation of the effectiveness of communication between psychologists and sales executives through personnel audit reports. Unpublished doctor's dissertation, Western Reserve Univer., 1955.)

We can provide further illustration from the Cornell data. The covariance between items in self-descriptions was factored, with the results shown in Table 4. This pattern is different from that of No. 3 (Table 3) in several respects. Notably, No. 3 overdifferentiates on all items. The first factor for No. 3 lumps openness and lack of receptiveness; these variables are divided among two factors in the criterion. In the criterion, being at ease (item 6) is positively related to openness. It is especially interesting that "feeling like the person being interviewed" is, for the group as a whole, positively correlated with being at ease; but for No. 3 these items are negatively correlated. When his expectancy is so discrepant from the facts, it is not surprising that No. 3 has poor accuracy.

RECOMMENDATIONS

Studies of perception may be concerned either with constant processes or with variable processes. When social perception is regarded (as in **1**, pp. 499–548) as a process of interpreting the expressive cues *O* presents, or of empathizing with him, the search is clearly for a variable process. The concept of an "intuitive" perception of *O*s which underlies much of the relevant research implies that *J* is reacting to the particular *O* as a stimulus, and ignores the fact that the perceptual response also depends on stereotypes in *J*'s mind (cf. Cattell, **5**). We have seen that the measures currently used are affected by both constant and reactive processes, and therefore cannot serve well to investigate either. As Crow states:

> The difficulty stems from failure to recognize that two meanings of predictive accuracy are involved. The use of the correlation scoring method [either $r_{x_{oi}y_{oij}}$ or $r_{x'_{oi}y'_{oij}}$] defines predictive accuracy as the ability to vary one's predictions as the actual situation varies. The difference score method defines predictive accuracy as the ability to approximate the actual situation. By the difference score method a subject is penalized for a systematic error in estimation of the magnitude of the actual situation. By the correlation method the subject is not so penalized. Conversely, a subject is penalized by the correlation method if, although he has approximated the actual situation, his predictions do not vary concomitantly with the actual scores. Each of these scoring methods has its advantages and disadvantages. The choice of which technique to use will depend on the purpose for which a study is conducted, although a second basis for choice depends upon the empirical relationship between two procedures (**10**, p. 57).

An argument can be presented for concentrating attention on constant processes, taking up interactions between *J* and *O* only after the constant processes characteristic of *J* are dependably measured. Constant processes in the perceiver have potentially great importance because they affect all his acts of perception. Individual differences in constant processes need to be measured dependably so that their influence can be discounted in studies of variable processes. Moreover, *identifying constant errors should permit training to eliminate such biases;* this may be the most effective way to improve the social perception of leaders, teachers, and diagnosticians.

Not all constant processes are of theoretical importance. We venture to suggest which components of social perception measures deserve attention, recognizing that the ultimate importance of the components depends on whether they relate to important criteria.

1. To some extent, the Elevation components (*E*, *DE*, *DEr*) reflect whether *J* interprets the words defining the scale in the same manner as *O*s do. It appears relatively unfruitful, therefore, as a source of information on his perception of *O*s (7; 8, p. 463). It should be separately measured where it is believed to have psychological significance, and otherwise eliminated from consideration. This is consistent with Postman's view:

> In experiments concerned with the determinants of perceptual selectivity, the contribution of verbal and motor response habits must be specifically evaluated and wherever possible held constant. The effects of the independent variables can then be evaluated against an empirical baseline defined by the response habits of the subjects (26, p. 26).

2. The Assumed Similarity measures reflect a general orientation toward Others. Perhaps the tendency to differentiate which these indices measure is a reaction shown only in the testing situation. But the fact that significant behavioral correlates have been found for Assumed Similarity (2, 4, 12, 13, 29) suggests that this is a generalized mental set influencing both test and nontest behavior. Investigators would do well, however, to consider Postman's conclusion that response dispositions can be established unambiguously only if they are measured by more than one type of response (26, p. 27).

Components of Assumed Similarity include Assumed Dispersion in Elevation, Assumed Dispersion over Items, Assumed Similarity in Elevation, and Assumed Self-Typicality. Further research is required to determine whether these should be measured separately or combined.

3. Stereotype Accuracy expresses how closely *J*'s implicit picture of the generalized Other agrees with reality. Differences of this sort are probably important. Attention should be given to the nature of *J*'s errors, as well as to the over-all magnitude of the component.

4. The *J*'s perceptual space, studied as a whole, includes not only information on his stereotype and his assumed dispersion, but also on the way in which he organizes the field of personality. This type of constant cognitive process appears to be a most important area for research.

5. The Differential Elevation Correlation and the Differential Accuracy Correlation are measures of *J*'s sensitivity to individual differences. These measures reflect his ability to interpret expressive behavior, or his ability in differential diagnosis. These are the only processes included in present measures of social perception which depend on *J*'s sensitivity to the particular *O*. The reliability of measures of this variable process has not been encouraging. But those who wish to study "empathy" or "social sensitivity" as it has usually been conceptualized should extract these correlational components from their measures.

Social perception research has been dominated by simple, operationally defined measures. Our analysis has shown that any such measure may combine and thereby conceal important variables, or may depend heavily on unwanted components. Only by careful subdivision of global measures can an investigator hope to know what he is dealing with. Our analysis makes especially clear that the investigator of social perception must develop more explicit theory regarding the constructs he intends to study, so that he can reduce his measures to the genuinely relevant components.

APPENDIX

To simplify this paper for the reader, we have placed our detailed mathematical argument here. In our notation, x_{oi} is the self-description of Other o on item $i(i=a, b, c, \cdots, k)$. y_{oij} is Judge j's description of o on i. We employ the customary notation for means: $\bar{x}_{o.}$ indicates the average of x_{oi} over all items, $\bar{x}_{.i}$ the average over Others, and $\bar{x}_{..}$, the grand mean over Others and items. Averages on y are defined similarly. $x_{oi}'=x_{oi}-\bar{x}_{o.}-\bar{x}_{.i}+\bar{x}_{..}$; that is, the score x_{oi} is transformed as a deviation from both item mean and Other mean. y_{oij}' is defined similarly.

Error in prediction may be measured by $|y_{oij}-x_{oi}|$. We shall, however, use the squared difference. This formula is easier to treat mathematically than the absolute difference, and will ordinarily give similar results. When all items are in a Yes-No form, so that the error on any prediction is 1 or 0, the two methods give identical results. Our measure has the important property of being invariant under orthogonal rotation of axes (8).

The Accuracy with which J perceives all Others is defined by

$$ACC_j^2=\frac{1}{kN}\sum_o\sum_i (y_{oij}-x_{oi})^2. \qquad [1a]$$

The following identity may be written:

$$\begin{aligned} y_{oij}-x_{oi}= {} & (\bar{y}_{..j}-\bar{x}_{..}) \\ & +[(\bar{y}_{o.j}-\bar{y}_{..j})-(\bar{x}_{o.}-\bar{x}_{..})] \\ & +[(\bar{y}_{.ij}-\bar{y}_{..j})-(\bar{x}_{.i}-\bar{x}_{..})] \\ & +(y_{oij}'-x_{oi}'). \qquad [2a] \end{aligned}$$

When we square and sum, cross-products drop out and we have the resolution of ACC_j^2 into components:

$$\begin{aligned} ACC_j^2= {} & \frac{1}{kN}\sum_o ACC_{oj}=(\bar{y}_{..j}-\bar{x}_{..})^2 \\ & +\frac{1}{N}\sum [(\bar{y}_{o.j}-\bar{y}_{..j})-(\bar{x}_{o.}-\bar{x}_{..})]^2 \\ & +\frac{1}{k}\sum [(\bar{y}_{.ij}-\bar{y}_{..j})-(\bar{x}_{.i}-\bar{x}_{..})]^2 \\ & +\frac{1}{kN}\sum_{oi}\sum (y_{oij}'-x_{oi}')^2. \qquad [3a] \end{aligned}$$

In order, these components are called E, DE, SA, and DA (squares being ignored except in equations). Each of the three latter terms may be rewritten as the variance of a difference. Equations 1 and 2 in the text indicate this form for DE and SA. We also have

$$DA_{ij}^2=\sigma_{y'_{oij}}^2+\sigma_{x'_{oi}}^2-2\sigma_{y'_{oij}}\sigma_{x'_{oi}}r_{y'_{oij}x'_{oi}}. \qquad [4a]$$

DA_{ij}^2, averaged over items yields DA_j^2. Each variance in the formula is taken over Others.

Some investigators have preferred to compute $DA_{oj}^2=\sum_i(x_{oi}'-y_{oi}')^2$. Summed over Others, this also yields DA_j^2. Subdivided, DA_{oj}^2 would yield a variance over items for the Other, and a "Q correlation" over items comparing predicted deviations for the Other with the actual deviations. This method of organizing the data is not recommended, because the correlations are critically dependent on the factorial content of the items employed and on the direction chosen as representing a high score on the item. In personality data, this direction is frequently arbitrary.

Assumed similarity is also defined in terms of a sum of squared differences:

$$\begin{aligned} AS_j^2= {} & \frac{1}{kN}\sum_o\sum_i (y_{oij}-x_{ij})^2=(\bar{y}_{..j}-\bar{x}_{.j})^2 \\ & +\sigma_{\bar{y}_{o.j}}^2 \\ & +\frac{1}{k}\sum [(\bar{y}_{.ij}-\bar{y}_{..j})-(\bar{x}_{ij}-\bar{x}_{.j})]^2 \\ & +\frac{1}{k}\sum_i \sigma_{y'_{oij}}^2. \qquad [5a] \end{aligned}$$

The components, in order, are AE, ADE, AST, and ADI. Only AST can be rewritten as a variance of differences:

$$AST^2=\sigma_{\bar{y}_{.ij}}^2+\sigma_{x_{ij}}^2-2\sigma_{x_{ij}}\sigma_{\bar{y}_{.ij}}r_{x_{ij}\bar{y}_{.ij}}. \qquad [6a]$$

This correlation is referred to as the Self-Typicality correlation (STr).

We assume that the goodness of predictions can be evaluated by the mean square error. Taking the derivative of each component of ACC_j^2, and setting that derivative equal to zero, we find that ACC becomes smaller and therefore prediction improves, when:

a. J has a typical response set (E approaches zero).

b. $\sigma_{\bar{y}_{.ij}}$ approaches $r_{\bar{x}_{.i}\bar{y}_{.ij}}\sigma_{\bar{x}_{.i}}$. Here the variance is over items. This means that $\sigma_{\bar{y}_{.ij}}$ should not exceed $\sigma_{\bar{x}_{.i}}$, and should be near zero if the Stereotype correlation is low. If this correlation is low, the more J differentiates among items, the poorer is his Accuracy.

c. $\sigma_{y'}$ approaches $r_{x_i'y_i'}\sigma_{x_i'}$, the variance being over Others. This means that $\sigma_{y'}$ should not exceed $\sigma_{x'}$, and should be near zero if the Differential correlation is low. This principle holds for accuracy of prediction on any single item, and for predicting elevation.

REFERENCES

1. ALLPORT, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
2. BIERI, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, 48, 61–66.

3. Bronfenbrenner, U., & Harding, J. Components and correlates of social sensitivity. Paper read at Amer. Psychol. Ass., New York, September, 1954.
4. Cass, Loretta K. Parent-child relationships and delinquency. *J. abnorm. soc. Psychol.*, 1952, **47**, 101–104.
5. Cattell, R. B. Measurement versus intuition in applied psychology. *J. Pers.*, 1937, **6**, 114–131.
6. Chowdhry, Kamla, & Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
7. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
8. Cronbach, L. J., & Gleser, Goldine C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
9. Cronbach, L. J., Hartmann, W., & Ehart, Mary E. *Investigation of the character and properties of assumed similarity measures.* Urbana, Ill: Bureau of Research and Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 7, Contract N6ori-07135.)
10. Crow, W. J. A methodological study of social perceptiveness. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
11. Dymond, Rosalind F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343–350.
12. Fiedler, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.
13. Fiedler, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–388.
14. Gage, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
15. Gage, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Pers.*, 1953, **22**, 128–141.
16. Gage, N. L., & Cronbach, L. J. *Conceptual and methodological problems in interpersonal perception.* Urbana, Ill.: Bureau of Educational Research, Univer. of Illinois, 1954 (Mimeo.)
17. Jones, E. E. Authoritarianism as a determinant of first-impression formation. *J. Pers.*, 1954, **23**, 107–127.
18. Kelly, G. A. *The psychology of personal constructs.* New York: Norton, in press.
19. Lorge, I., & Diamond, Lorraine K. The value of information to good and poor judges of item difficulty. *Educ. psychol. Measmt*, 1954, **14**, 29–33.
20. Melton, R. S. A comparison of clinical and actuarial methods of prediction with an assessment of the relative accuracy of different clinicians. Unpublished doctor's dissertation, Univer. of Minnesota, 1953.
21. Melton, R. S. A study of the relative accuracy of counselor judgments and actuarial predictions. *Amer. Psychologist*, 1954, **9**, 429–430. (Abstract)
22. Moore, J. V. Factor analytic comparisons of superior and subordinate ratings of the same NCO supervisors. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1953, No. 53-24.
23. Newcomb, T. M. An experiment designed to test the validity of a rating technique. *J. educ. Psychol.*, 1937, **22**, 279–289.
24. Norman, R. D., & Ainsworth, Patricia. The relationships among projection, empathy, reality, and adjustment, operationally defined. *J. consult. Psychol.*, 1954, **18**, 53–58.
25. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
26. Postman, L. Perception, motivation, and behavior. *J. Pers.*, 1953, **22**, 17–31.
27. Sarbin, T. R. The relative accuracy of clinical and statistical predictions of academic success. Unpublished doctor's dissertation, Ohio State Univer., 1941.
28. Sarbin, T. R. A contribution to the study of actuarial and individual methods of prediction. *Amer. J. Sociol.*, 1943, **48**, 593–602.
29. Sopchak, A. L. Parental "identification" and "tendency toward disorders" as measured by the Minnesota Multiphasic Personality Inventory. *J. abnorm. soc. Psychol.*, 1952, **47**, 159–165.
30. Soskin, W. F. Frames of reference in personality assessment. *J. clin. Psychol.*, 1954, **10**, 107–114.
31. Steiner, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol.*, 1954, **49**, 349–354.
32. Stone, G. C., & Leavitt, G. S. Generality of accuracy in perceiving standard persons. Paper read at Midwest. Psychol. Ass., Columbus, Ohio, May, 1954.
33. Talland, G. A. The assessment of group opinion by leaders, and their influence on its formation. *J. abnorm. soc. Psychol.*, 1954, **49**, 431–434.
34. Wickman, E. K. *Children's behavior and teachers' attitudes.* New York: Commonwealth Fund, 1928.

Received May 5, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 3, 1955

ANTECEDENT PROBABILITY AND THE EFFICIENCY OF PSYCHOMETRIC SIGNS, PATTERNS, OR CUTTING SCORES

PAUL E. MEEHL
University of Minnesota

AND ALBERT ROSEN
VA Hospital, Minneapolis, and University of Minnesota[1]

In clinical practice, psychologists frequently participate in the making of vital decisions concerning the classification, treatment, prognosis, and disposition of individuals. In their attempts to increase the number of correct classifications and predictions, psychologists have developed and applied many psychometric devices, such as patterns of test responses as well as cutting scores for scales, indices, and sign lists. Since diagnostic and prognostic statements can often be made with a high degree of accuracy purely on the basis of actuarial or experience tables (referred to hereinafter as *base rates*), a psychometric device, to be efficient, must make possible a greater number of correct decisions than could be made in terms of the base rates alone.

The efficiency of the great majority of psychometric devices reported in the clinical psychology literature is difficult or impossible to evaluate for the following reasons:

a. Base rates are virtually never reported. It is, therefore, difficult to determine whether or not a given device results in a greater number of correct decisions than would be possible solely on the basis of the rates from previous experience. When, however, the base rates can be estimated, the reported claims of efficiency of psychometric instruments are often seen to be without foundation.

b. In most reports, the distribution data provided are insufficient for the evaluation of the probable efficiency of the device in other settings where the base rates are markedly different. Moreover, the samples are almost always too small for the determination of optimal cutting lines for various decisions.

c. Most psychometric devices are reported without cross-validation data. If a psychometric instrument is applied solely to the criterion groups from which it was developed, its reported validity and efficiency are likely to be spuriously high, especially if the criterion groups are small.

d. There is often a lack of clarity concerning the type of population in which a psychometric device can be effectively applied.

e. Results are frequently reported only in terms of significance tests for differences between groups rather than in terms of the number of correct decisions for individuals within the groups.

The purposes of this paper are to examine current methodology in studies of predictive and concurrent validity (1), and to present some methods for the evaluation of the efficiency of psychometric devices as well as for the improvement in the interpretations made from such devices. Actual studies reported in the literature will be used for illustration wherever possible. It should be emphasized that these particular illustrative studies of common practices were chosen simply because they contained more complete data than are commonly reported, and were available in fairly recent publications.

---

[1] From the Neuropsychiatric Service, VA Hospital, Minneapolis, Minnesota, and the Divisions of Psychiatry and Clinical Psychology of the University of Minnesota Medical School. The senior author carried on his part of this work in connection with his appointment to the Minnesota Center for the Philosophy of Science.

IMPORTANCE OF BASE RATES

Danielson and Clark (4) have reported on the construction and application of a personality inventory which was devised for use in military induction stations as an aid in detecting those men who would not complete basic training because of psychiatric disability or AWOL recidivism. One serious defect in their article is that it reports cutting lines which have not been cross validated. Danielson and Clark state that inductees were administered the Fort Ord Inventory within two days after induction into the Army, and that all of these men were allowed to undergo basic training regardless of their test scores.

Two samples (among others) of these inductees were selected for the study of predictive validity: (*a*) A group of 415 men who had made a good adjustment (Good Adjustment Group), and (*b*) a group of 89 men who were unable to complete basic training and who were sufficiently disturbed to warrant a recommendation for discharge by a psychiatrist (Poor Adjustment Group). The authors state that "the most important task of a test designed to screen out misfits is the detection of the (latter) group" (4, p. 139). The authors found that their most effective scale for this differentiation picked up, at a given cutting point, 55% of the Poor Adjustment Group (valid positives) and 19% of the Good Adjustment Group (false positives). The overlap between these two groups would undoubtedly have been greater if the cutting line had been cross validated on a random sample from the *entire population* of inductees, but for the purposes of the present discussion, let us assume that the results were obtained from cross-validation groups. There is no mention of the percentage of all inductees who fall into the Poor Adjustment Group, but a rough estimate will be adequate for the present discussion. Suppose that in their population of soldiers, as many as 5% make a poor adjustment and 95% make a good adjustment. The results for 10,000 cases would be as depicted in Table 1.

TABLE 1

NUMBER OF INDUCTEES IN THE POOR ADJUSTMENT AND GOOD ADJUSTMENT GROUPS DETECTED BY A SCREENING INVENTORY

(55% valid positives; 19% false positives)

Predicted Adjustment	Actual Adjustment				Total Predicted
	Poor		Good		
	No.	%	No.	%	
Poor	275	55	1,805	19	2,080
Good	225	45	7,695	81	7,920
Total actual	500	100	9,500	100	10,000

Efficiency in detecting poor adjustment cases. The efficiency of the scale can be evaluated in several ways. From the data in Table 1 it can be seen that if the cutting line given by the authors were used at Fort Ord, the scale could not be used directly to "screen out misfits." If all those predicted by the scale to make a poor adjustment were screened out, the number of false positives would be extremely high. Among the 10,000 potential inductees, 2080 would be predicted to make a poor adjustment. Of these 2080, only 275, or 13%, would actually make a poor adjustment, whereas the decisions for 1805 men, or 87% of those screened out, would be incorrect.

Efficiency in prediction for all cases. If a prediction were made for every man on the basis of the cutting line given for the test, 275+7695, or 7970, out of 10,000 decisions would be correct. Without the test, however, every man would be predicted to make a good adjustment, and 9500 of the predictions would be correct. Thus, use of the test has yielded a drop from 95% to 79.7% in the total number of correct decisions.

Efficiency in detecting good adjustment cases. There is one kind of decision in which the Inventory can improve on the base rates, however. If only those men are accepted who are predicted by the Inventory to make a good adjustment, 7920 will be selected, and the outcome of 7695 of the 7920, or 97%, will be predicted correctly. This is a 2% increase in hits among predictions of "success." The decision as to whether or not the scale improves on the base rates sufficiently to warrant its use will depend on the cost of administering the testing program, the administrative feasibility of rejecting 21% of the men who passed the psychiatric screening, the cost to the Army of training the 225 maladaptive recruits, and the intangible human costs involved in psychiatric breakdown.

Populations to which the scale is applied. In the evaluation of the efficiency of any psychometric instrument, careful consideration must be given to the types of populations to which the device is to be applied. Danielson and Clark have stated that "since the final decision as to disposition is made by the psychiatrist, the test should be classified as a screening adjunct" (4, p. 138). This statement needs clarification, however, for the efficiency of the scale can vary markedly according to the different ways in which it might be used as an adjunct.

It will be noted that the test was administered to men who were already in the Army, and not to men being examined for induction. The reported validation data apply, therefore, specifically to the population of *recent inductees*. The results might have been somewhat different if the population tested consisted of *potential inductees*. For the sake of illustration, however, let us assume that there is no difference in the test results of the two populations.

An induction station psychiatrist can use the scale cutting score in one or more of the following ways, i.e., he can apply the scale results to a variety of populations. (*a*) The psychiatrist's final decision to accept or reject a potential inductee may be based on both the test score and his usual interview procedure. The population to which the test scores are applied is, therefore, *potential inductees interviewed by the usual procedures for whom no decision was made.* (*b*) He may evaluate the potential inductee according to his usual procedures, and then consult the test score *only if* the tentative decision is to reject. That is, a decision to accept is final. The population to which the test scores are applied is *potential inductees tentatively rejected by the usual interview procedures.* (*c*) An alternative procedure is for the psychiatrist to consult the test score only if the tentative decision is to accept, the population being *potential inductees tentatively accepted by the usual interview procedures.* The decision to reject is final. (*d*) Probably the commonest proposal for the use of tests as screening adjuncts is that the more skilled and costly psychiatric evaluation should be made only upon the test positives, i.e., inductees classified by the test as good risks are not interviewed, or are subjected only to a very short and superficial interview. Here the population is *all potential inductees*, the test being used to make either a *final* decision to "accept" or a decision to "examine."

Among these different procedures, how is the psychiatrist to achieve maximum effectiveness in using the test as an adjunct? There is no answer to this question from the available data, but it can be stated definitely that the data reported by Danielson and Clark apply only to the third procedure described above. The test results are based on a selected group of men *accepted* for induction and not on a random sample of potential inductees. If the scale is used in any other way than the third procedure mentioned above, the results may be considerably inferior to those reported, and, thus, to the use of the base rates without the test.[2]

The principles discussed thus far, although illustrated by a single study, can be generalized to any study of predictive or concurrent validity. It can be seen that many considerations are involved in determining the efficiency of a scale at a given cutting score, especially the base rates of the subclasses within the population to which the psychometric device is to be applied. In a subsequent portion of this paper, methods will be presented for determining cutting points for maximizing the efficiency of the different types of decisions which are made with psychometric devices.

Another study will be utilized to illustrate the importance of an explicit statement of the base rates of population subgroups to be tested with a given device. Employing an interesting configural approach, Thiesen (**18**) discovered five Rorschach patterns, each of which differentiated well between 60 schizophrenic adult patients and a sample of 157 gainfully employed adults. The best differentiator, considering individual patterns or number of patterns, was Pattern A, which was found in 20% of the patients' records and in only .6% of the records of normals. Thiesen concludes that if these patterns stand the test of cross validation, they might have "clinical usefulness" in early detection of a schizophrenic process or as an aid to determining the gravity of an initial psychotic episode (**18**, p. 369). If by "clinical usefulness" is meant efficiency in a clinic or hospital for the diagnosis of schizophrenia, it is necessary to demonstrate that the patterns differentiate a higher percentage of schizophrenic patients from *other diagnostic groups* than could be correctly classified without any test at all, i.e., solely on the basis of the rates of various diagnoses in any given hospital. If a test is to be used in differential diagnosis among psychiatric patients, evidence of its efficiency for this function cannot be established solely on the basis of discrimination of diagnostic groups from normals. If by "clinical usefulness" Thiesen means that his data indicate that the patterns might be used to detect an early schizophrenic process among nonhospitalized gainfully employed adults, he would do better to discard his patterns and use the base rates, as can be seen from the following data.

Taulbee and Sisson (**17**) cross validated Thiesen's patterns on schizophrenic patient and normal samples, and found that Pattern A was the best discriminator. Among patients,

---

[2] Goodman (**8**) has discussed this same problem with reference to the supplementary use of an index for the prediction of parole violation.

8.1% demonstrated this pattern and among normals, none had this pattern. There are approximately 60 million gainfully employed adults in this country, and it has been estimated that the rate of schizophrenia in the general population is approximately .85% (**2**, p. 558). The results for Pattern A among a population of 10,000 gainfully employed adults would be as shown in Table 2. In order to detect 7 schizophrenics, it would be necessary to test 10,000 individuals.

TABLE 2

NUMBER OF PERSONS CLASSIFIED AS SCHIZOPHRENIC AND NORMAL BY A TEST PATTERN AMONG A POPULATION OF GAINFULLY EMPLOYED ADULTS

(8.1% valid positives; 0.0% false positives)

Classification by Test	Criterion Classification				Total Classified by Test
	Schizophrenia		Normal		
	No.	%	No.	%	
Schizophrenia	7	8.1	0	0	7
Normal	78	91.9	9,915	100	9,993
Total in class	81	100	9,915	100	10,000

In the Neurology service of a hospital a psychometric scale is used which is designed to differentiate between patients with psychogenic and organic low back pain (**9**). At a given cutting point, this scale was found to classify each group with approximately 70% effectiveness upon cross validation, i.e., 70% of cases with no organic findings scored above an optimal cutting score, and 70% of surgically verified organic cases scored below this line. Assume that 90% of all patients in the Neurology service with a primary complaint of low back pain are in fact "organic." Without any scale at all the psychologist can say every case is organic, and be right 90% of the time. With the scale the results would be as shown in Section A of Table 3. Of 10 psychogenic cases, 7 score above the line; of 90 organic cases, 63 score below the cutting line. If every case above the line is called psychogenic, only 7 of 34 will be classified correctly or about 21%. Nobody wants to be right only one out of five times in this type of situation, so that it is obvious that it would be imprudent to call a patient psychogenic on the basis of this scale. Radically different results occur in prediction for cases below the cutting line. Of 66 cases 63, or 95%, are correctly classified as organic. Now the psychologist has increased his diagnostic hits from 90 to 95% on the condition that he labels only cases falling below the line, and ignores the 34% scoring above the line.

TABLE 3

NUMBER OF PATIENTS CLASSIFIED AS PSYCHOGENIC AND ORGANIC ON A LOW BACK PAIN SCALE WHICH CLASSIFIES CORRECTLY 70% OF PSYCHOGENIC AND ORGANIC CASES

Classification by Scale	Actual Diagnosis		Total Classified by Scale
	Psychogenic	Organic	
A. Base Rates in Population Tested: 90% Organic; 10% Psychogenic			
Psychogenic	7	27	34
Organic	3	63	66
Total diagnosed	10	90	100
B. Base Rates in Population Tested: 90% Psychogenic; 10% Organic			
Psychogenic	63	3	66
Organic	27	7	34
Total diagnosed	90	10	100

In actual practice, the psychologist may not, and most likely will not, test every low back pain case. Probably those referred for testing will be a select group, i.e., those who the neurologist believes are psychogenic because neurological findings are minimal or absent. This fact changes the population from "all patients in Neurology with a primary complaint of low back pain," to "all patients in Neurology with a primary complaint of low back pain *who are referred for testing*." Suppose that a a study of past diagnoses indicated that of patients with minimal or absent findings, 90% were diagnosed as psychogenic and 10% as organic. Section B of Table 3 gives an entirely different picture of the effectiveness of the low back pain scale, and new limitations on interpretation are necessary. Now the scale correctly classifies 95% of all cases above the line as psychogenic (63 of 66), and is correct in only 21% of all cases below the line (7 of 34). In this practical situation the psychologist would be wise to refrain from interpreting a low score.

From the above illustrations it can be seen that the psychologist in interpreting a test and in evaluating its effectiveness must be very much aware of the population and its subclasses and the base rates of the behavior or event with which he is dealing at any given time.

It may be objected that no clinician relies on just one scale but would diagnose on the basis of a configuration of impressions from several tests, clinical data and history. We must, therefore, emphasize that the preceding single-scale examples were presented for simplicity only, but that the main point is not dependent upon this "atomism." *Any complex configurational procedure in any number of variables, psychometric or otherwise, eventuates in a decision.* Those decisions have a certain objective success rate in criterion case identification; and for present purposes we simply treat the decision function, whatever its components and complexity may be, as a single variable. It should be remembered that the literature does not present us with cross-validated methods having hit rates much above those we have chosen as examples, regardless of how complex or configural the methods used. So that even if the clinician approximates an extremely complex configural function "in his head" before classifying the patient, for purposes of the present problem this complex function is treated as the scale. In connection with the more general "philosophy" of clinical decision making see Bross (**3**) and Meehl (**12**).

APPLICATIONS OF BAYES' THEOREM

Many readers will recognize the preceding numerical examples as essentially involving a principle of elementary probability theory, the so-called "Bayes' Theorem." While it has come in for some opprobrium on account of its connection with certain pre-Fisherian fallacies in statistical inference, as an algebraic statement the theorem has, of course, nothing intrinsically wrong with it and it does apply in the present case. One form of it may be stated as follows:

If there are k antecedent conditions under which an event of a given kind may occur, these conditions having the antecedent probabilities P_1, P_2, $\cdots$, P_k of being realized, and the probability of the event upon each of them is p_1, p_2, p_3, $\cdots$, p_k; then, given that the event is observed to occur, the probability that it arose on the basis of a specified one, say j, of the antecedent conditions is given by

$$P_{j(o)}=\frac{P_j p_j}{\sum_{i=1}^{k} P_i p_i}.$$

The usual illustration is the case of drawing marbles from an urn. Suppose we have two urns, and the urn-selection procedure is such that the probability of our choosing the first urn is 1/10 and the second 9/10. Assume that 70% of the marbles in the first urn are black, and 40% of those in the second urn are black. I now (blindfolded) "choose" an urn and then, from it, I choose a marble. The marble turns out to be black. What is the probability that I drew from the first urn?

$$P_1=.10 \qquad P_2=.90$$

$$p_1=.70 \qquad p_2=.40$$

Then

$$P_{1(b)}=\frac{(.10)(.70)}{(.10)(.70)+(.90)(.40)}=.163.$$

If I make a practice of inferring under such circumstances that an observed black marble arose from the first urn, I shall be correct in such judgments, in the long run, only 16.3% of the time. Note, however, that the "test item" or "sign" *black marble* is correctly "scored" in favor of Urn No. 1, since there is a 30% difference in black marble rate between it and Urn No. 2. But this considerable disparity in symptom rate is overcome by the very low base rate ("antecedent probability of choosing from the first urn"), so that inference to first-urn origin of black marbles will actually be wrong some 84 times in 100. In the clinical analogue, the urns are identified with the subpopulations of patients to be discriminated (their antecedent probabilities being equated to their base rates in the population to be examined), and the black marbles are test results of a certain ("positive") kind. The proportion of black marbles in one urn is the valid positive rate, and in the other is the false positive rate. Inspection and suitable manipulations of the formula for the common two-category case, viz.,

$$P_{(o)}=\frac{Pp_1}{Pp_1+Qp_2}$$

$P_{d(o)}$ = Probability that an individual is diseased, given that his observed test score is positive
P = Base rate of actual positives in the population examined
$P+Q=1$
p_1 = Proportion of diseased identified by test ("valid positive" rate)
$q_1=1-p_1$
p_2 = Proportion of nondiseased misidentified by test as being diseased ("false positive" rate)
$q_2=1-p_2$

yields several useful statements. Note that in what follows we are operating entirely with exact population parameter values; i.e., sampling errors are not responsible for the dangers and restrictions set forth. See Table 4.

1. In order for a positive diagnostic assertion to be "more likely true than false," the ratio of the positive to the negative base rates in the examined population must exceed the ratio of the false positive rate to the valid positive rate. That is,

$$\frac{P}{Q}>\frac{p_2}{p_1}.$$

If this condition is not met, the attribution of pathology on the basis of the test is more probably in error than correct, *even though the sign being used is valid* (i.e., $p_1 \neq p_2$).

TABLE 4

Definition of Symbols

Diagnosis from Test	Actual Diagnosis	
	Positive	Negative
Positive	p_1 Valid positive rate (Proportion of positives called positive)	p_2 False positive rate (Proportion of negatives called positive)
Negative	q_1 False negative rate (Proportion of positives called negative)	q_2 Valid negative rate (Proportion of negatives called negative)
Total with actual diagnosis	$p_1+q_1=1.0$ (Total positives)	$p_2+q_2=1.0$ (Total negatives)

Note.—For simplicity, the term "diagnosis" is used to denote the classification of any kind of pathology, behavior, or event being studied, or to denote "outcome" if a test is used for prediction. Since horizontal addition (e.g., p_1+p_2) is meaningless in ignorance of the base rates, there is no symbol or marginal total for these sums. *All values are parameter values.*

Example: If a certain cutting score identifies 80% of patients with organic brain damage (high scores being indicative of damage) but is also exceeded by 15% of the nondamaged sent for evaluation, in order for the psychometric decision "brain damage present" to be more often true than false, the ratio of actually brain-damaged to nondamaged cases among all seen for testing must be at least one to five (.19).

Piotrowski has recommended that the presence of 5 or more Rorschach signs among 10 "organic" signs is an efficient indicator of brain damage. Dorken and Kral (5), in cross validating Piotrowski's index, found that 63% of organics and 30% of a mixed, nonorganic, psychiatric patient group had Rorschachs with 5 or more signs. Thus, our estimate of $p_2/p_1 = .30/.63 = .48$, and in order for the decision "brain damage present" to be correct more than one-half the time, the proportion of positives (P) in a given population must exceed .33 (i.e., $P/Q > .33/.67$). Since few clinical populations requiring this clinical decision would have such a high rate of brain damage, especially among psychiatric patients, the particular cutting score advocated by Piotrowski will produce an excessive number of false positives, and the positive diagnosis will be more often wrong than right. Inasmuch as the base rates for any given behavior or pathology differ from one clinical setting to another, *an inflexible cutting score should not be advocated for any psychometric device.* This statement applies generally—thus, to indices recommended for such diverse purposes as the classification or detection of deterioration, specific symptoms, "traits," neuroticism, sexual aberration, dissimulation, suicide risk, and the like. When P is small, it may be advisable to explore the possibility of dealing with a restricted population within which the base rate of the attribute being tested is higher. This approach is discussed in an article by Rosen (**14**) on the detection of suicidal patients in which it is suggested that an attempt might be made to apply an index to subpopulations with higher suicide rates.

2. If the base rates are equal, the probability of a positive diagnosis being correct is the ratio of valid positive rate to the sum of valid and false positive rates. That is,

$$p_{d(o)} = \frac{p_1}{p_1 + p_2} \quad \text{if} \quad P = Q = \tfrac{1}{2}.$$

Example: If our population is evenly divided between neurotic and psychotic patients the condition for being "probably right" in diagnosing psychosis by a certain method is simply that the psychotics exhibit the pattern in question more frequently than the neurotics. This is the intuitively obvious special case; it is often misgeneralized to justify use of the test in those cases where base-rate asymmetry ($P \neq Q$) counteracts the ($p_1 - p_2$) discrepancy, leading to the paradoxical consequence that *deciding on the basis of more information can actually worsen the chances of a correct decision.* The apparent absurdity of such an idea has often misled psychologists into behaving as though the establishment of "validity" or "discrimination," i.e., that $p_1 \neq p_2$, indicates that a procedure should be used in decision making.

Example: A certain test is used to select those who will continue in outpatient psychotherapy (positives). It correctly identifies 75% of these good cases but the same cutting score picks up 40% of the poor risks who subsequently terminate against advice. Suppose that in the past experience of the clinic 50% of the patients terminated therapy prematurely. Correct selection of patients can be made with the given cutting score on the test 65% of the time, since $p_1/(p_1+p_2) = .75/(.75+.40) = .65$. It can be seen that the efficiency of the test would be exaggerated if the base rate for continuation in therapy were actually .70, but the efficiency were evaluated solely on the basis of a research study containing equal groups of continuers and noncontinuers, i.e., if it were assumed that $P = .50$.

3. In order for the hits in the entire population which is under consideration to be increased by use of the test, the base rate of the more numerous class (called here positive) must be less than the ratio of the valid negative rate to the sum of valid negative and false negative rates. That is, unless

$$P < \frac{q_2}{q_1 + q_2},$$

the making of decisions on the basis of the test will have an adverse effect. An alternative expression is that $(P/Q) < (q_2/q_1)$ when $P > Q$, i.e., the ratio of the larger to the smaller class must be less than the ratio of the valid negative rate to the false negative rate. When $P < Q$, the conditions for the test to improve upon the base rates are:

$$Q < \frac{p_1}{p_1 + p_2}$$

and

$$\frac{Q}{P} < \frac{p_1}{p_2}.$$

Rotter, Rafferty, and Lotsof (**15**) have reported the scores on a sentence completion test for a group of 33 "maladjusted" and 33 "adjusted" girls. They report that the use of a specified cutting score (not cross validated) will result in the correct classification of 85% of the maladjusted girls and the incorrect classification of only 15% of the adjusted girls. It is impossible to evaluate adequately the efficiency of the test unless one knows the base rates of maladjustment (P) and adjustment (Q) for the population of high school girls, although there would be general agreement that $Q > P$. Since $p_1/(p_1+p_2) = .85/(.85+.15) = .85$, the over-all hits in diagnosis with the test will not improve on classification based solely on the base rates unless the proportion of adjusted girls is less than .85. Because the reported effectiveness of the test is spuriously high, the proportion of adjusted girls would no doubt have to be considerably less than .85. Unless there is good reason to believe that the base rates are similar from one setting to another, it is impossible to determine the efficiency of a test such as Rotter's when the criterion is based on ratings unless one replicates his research, including the criterion ratings, with a representative sample of each new population.

4. In altering a sign, improving a scale, or shifting a cutting score, the increment in valid positives per increment in valid positive *rate* is proportional to the positive base rate; and analogously, the increment in valid negatives per increment in valid negative *rate* is proportional to the negative base rate. That is, if we alter a sign the net improvement in over-all hit rate is

$$H'_T - H_T = \Delta p_1 P + \Delta q_2 Q,$$

where H_T = original proportion of hits (over-all) and H'_T = new proportion of hits (over-all).

5. A corollary of this is that altering a sign or shifting a cut will improve our decision making if, and only if, the ratio of *improvement* Δp_1 in valid positive rate to *worsening* Δp_2 in false negative rate exceeds the ratio of actual negatives to positives in the population.

$$\frac{\Delta p_1}{\Delta p_2} > \frac{Q}{P}\,.$$

Example: Suppose we improve the intrinsic validity of a certain "schizophrenic index" so that it now detects 20% more schizophrenics than it formerly did, at the expense of only a 5% increase in the false positive rate. This surely looks encouraging. We are, however, working with an outpatient clientele only 1/10th of whom are actually schizophrenic. Then, since

$$\Delta p_1 = .20 \qquad P = .10$$

$$\Delta p_2 = .05 \qquad Q = .90$$

applying the formula we see that

$$\frac{.20}{.05} \ngtr \frac{.90}{.10}$$

i.e., the required inequality does not hold, and the routine use of this "improved" index will result in an increase in the proportion of erroneous diagnostic decisions.

In the case of any pair of unimodal distributions, this corresponds to the principle that the optimal cut lies at the intersection of the two distribution envelopes (**11**, pp. 271–272).

MANIPULATION OF CUTTING LINES FOR DIFFERENT DECISIONS

For any given psychometric device, no one cutting line is maximally efficient for clinical settings in which the base rates of the criterion groups in the population are different. Furthermore, different cutting lines may be necessary for various decisions within the same population. In this section, methods are presented for manipulating the cutting line of any instrument in order to maximize the efficiency of a device in the making of several kinds of decisions. Reference should be made to the scheme presented in Table 5 for understanding of the discussion which follows. This scheme and the methods for manipulating cutting lines are derived from Duncan, Ohlin, Reiss, and Stanton (**6**).

A study in the prediction of juvenile delinquency by Glueck and Glueck (**7**) will be used for illustration. Scores on a prediction index for 451 delinquents and 439 nondelinquents (**7**, p. 261) are listed in Table 6. If the Gluecks' index is to be used in a population with a given juvenile delinquency rate, cutting lines can be established to maximize the efficiency of the index for several de-

TABLE 5

SYMBOLS TO BE USED IN EVALUATING THE EFFICIENCY OF A PSYCHOMETRIC DEVICE IN CLASSIFICATION OR PREDICTION

Diagnosis from Test	Actual Diagnosis: Positive	Actual Diagnosis: Negative	Total Diagnosed from Test
Positive	NPp_1 (Number of valid positives)	NQp_2 (Number of false positives)	NPp_1+NQp_2 (Number of test positives)
Negative	NPq_1 (Number of false negatives)	NQq_2 (Number of valid negatives)	NPq_1+NQq_2 (Number of test negatives)
Total with actual diagnosis	NP (Number of actual positives)	NQ (Number of actual negatives)	N (Total number of cases)

Note.—For simplicity, the term "diagnosis" is used to denote the classification of any kind of pathology, behavior, or event studied, or to denote "outcome" if a test is used for prediction. "Number" means *absolute frequency*, not *rate* or probability.

cisions. In the following illustration, a delinquency rate of .20 will be used. From the data in Table 6, optimal cutting lines will be determined for maximizing the proportion of correct predictions, or hits, for all cases (H_T), and for maximizing the proportion of hits (H_P) among those called delinquent (positives) by the index.

In the first three columns of Table 6, "*f*" denotes the number of delinquents scoring in each class interval, "*cf*" represents the cumulative frequency of delinquents scoring above each class interval (e.g., 265 score above 299), and p_1 represents the proportion of the total group of 451 delinquents scoring above each class interval. Columns 4, 5, and 6 present the same kind of data for the 439 nondelinquents.

Maximizing the number of correct predictions or classifications for all cases. The proportion of correct predictions or classifications (H_T) for any given cutting line is given by the formula, $H_T = Pp_1 + Qq_2$. Thus, in

TABLE 6

PREDICTION INDEX SCORES FOR JUVENILE DELINQUENTS AND NONDELINQUENTS AND OTHER STATISTICS FOR DETERMINING OPTIMAL CUTTING LINES FOR CERTAIN DECISIONS IN A POPULATION WITH A DELINQUENCY RATE OF .20*

Prediction Index Score	Delinquents			Nondelinquents									
			$\frac{\text{cf}}{451}$			$\frac{\text{cf}}{439}$	$1-p_2$	$.2p_1$	$.8p_2$	$.8q_2$	Pp_1+Qq_2	Pp_1+Qp_2	$\frac{Pp_1}{R_P}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	f	*cf*	p_1	*f*	*cf*	p_2	q_2	Pp_1	Qp_2	Qq_2	H_T	R_P	H_P
400+	51	51	.1131	1	1	.0023	.9977	.0226	.0018	.7982	.821	.024	.926
350–399	73	124	.2749	8	9	.0205	.9795	.0550	.0164	.7836	.839	.071	.770
300–349	141	265	.5876	23	32	.0729	.9271	.1175	.0583	.7417	.859	.176	.668
250–299	122	387	.8581	70	102	.2323	.7677	.1716	.1858	.6142	.786	.357	.480
200–249	40	427	.9468	68	170	.3872	.6128	.1894	.3098	.4902	.680	.499	.379
150–199	19	446	.9889	102	272	.6196	.3804	.1978	.4957	.3043	.502	.694	.285
<150	5	451	1.0000	167	439	1.0000	.0000	.2000	.8000	.0000	.200	1.000	.200

* Frequencies in columns 1 and 4 are from Glueck and Glueck (7, p. 261).

column 11 of Table 6, labelled H_T, it can be seen that the best cutting line for this decision would be between 299 and 300, for 85.9% of all predictions would be correct if those above the line were predicted to become delinquent and all those below the line nondelinquent. Any other cutting line would result in a smaller proportion of correct predictions, and, in fact, any cutting line set lower than this point would make the index inferior to the use of the base rates, for if all cases were predicted to be nondelinquent, the total proportion of hits would be .80.

Maximizing the number of correct predictions or classifications for positives. The primary use of a prediction device may be for *selection* of (*a*) students who will succeed in a training program, (*b*) applicants who will succeed in a certain job, (*c*) patients who will benefit from a certain type of therapy, etc. In the present illustration, the index would most likely be used for detection of those who are likely to become delinquents. Thus, the aim might be to maximize the number of hits only within the group predicted by the index to become delinquents (predicted positives $= NPp_1 + NQp_2$). The proportion of correct predictions for this group by the use of different cutting lines is given in column 13, labelled H_P. Thus, if a cutting line is set between 399 and 400, one will be correct over 92 times in 100 if predictions are made *only* for persons scoring above the cutting line. The formula for determining the efficiency of the test when only positive predictions are made is $H_P = Pp_1/(Pp_1 + Qp_2)$.

One has to pay a price for achieving a very high level of accuracy with the index. Since the problem is to select potential delinquents so that some sort of therapy can be attempted, the proportion of this selected group in the total sample may be considered as a selection ratio. The selection ratio for positives is $R_P = Pp_1 + Qp_2$, that is, predictions are made only for those above the cutting line. The selection ratio for each posssible cutting line is shown in column 12 of Table 6, labelled R_P. It can be seen that to obtain maximum accuracy in selection of delinquents (92.6%), predictions can be made for only 2.4% of the population. For other cutting lines, the accuracy of selection and the corresponding selection ratios are given in Table 6. The worker applying the index must use his own judgment in deciding upon the level of accuracy and the selection ratio desired.

Maximizing the number of correct predictions or classifications for negatives. In some selection problems, the goal is the selection of negatives rather than positives. Then, the proportion of hits among all predicted negative for any given cutting line is $H_N = Qq_2/(Qq_2 + Pq_1)$, and the selection ratio for negatives is $R_N = Pq_1 + Qq_2$.

In all of the above manipulations of cutting lines, it is essential that there be a large number of cases. Otherwise, the percentages about any given cutting line would be so unstable that very dissimilar results would be obtained on new samples. For most studies in clinical psychology, therefore, it would be necessary to establish cutting lines according to the decisions and methods discussed above, and then to cross validate a specific cutting line on new samples.

The amount of shrinkage to be expected in the cross validation of cutting lines cannot be determined until a thorough mathematical and statistical study of the subject is made. It may be found that when criterion distributions are approxi-

TABLE 7

PERCENTAGE OF DELINQUENTS (D) AND NONDELINQUENTS (ND) IN EACH PREDICTION INDEX SCORE INTERVAL IN A POPULATION IN WHICH THE DELINQUENCY RATE IS .20*

Prediction Index Score Interval	No. of D	No. of ND	Total of D and ND	% of D in Score Interval	% of ND in Score Interval	% of D and ND in Score Interval
400+	51	4	55	92.7	7.3	100
350–399	73	33	106	68.9	31.1	100
300–349	141	95	236	59.7	40.3	100
250–299	122	288	410	29.8	70.2	100
200–249	40	279	319	12.5	87.5	100
150–199	19	419	438	4.3	95.7	100
<150	5	686	691	.7	99.3	100
Total	451	1804	2255			

* Modification of Table XX-2, p. 261, from Glueck and Glueck (7).

mately normal and large, cutting lines should be established in terms of the normal probability table rather than on the basis of the observed p and q values found in the samples. In a later section dealing with the selection ratio we shall see that it is sometimes the best procedure to select all individuals falling above a certain cutting line and to select the others needed to reach the selection ratio by choosing at random below the line; or in other cases to establish several different cuts defining *ranges* within which one or the opposite decision should be made.

Decisions based on score intervals rather than cutting lines. The Gluecks' data can be used to illustrate another approach to psychometric classification and prediction when scores for large samples are available with a relatively large number of cases in each score interval. In Table 7 are listed frequencies of delinquents and nondelinquents for prediction index score intervals. The frequencies for delinquents are the same as those in Table 6, whereas those for nondelinquents have been corrected for a base rate of .20 by multiplying each frequency in column 4 of Table 6[3] by

$$4.11=\frac{(.80)\ (459)}{(.20)\ (431)}.$$

Table 7 indicates the proportion of delinquents and nondelinquents among all juveniles who fall within a given score interval when the base rate of delinquency is .20. It can be predicted that of those scoring 400 or more, 92.7% will become delinquent, of those scoring between 350 and 399, 68.9% will be delinquent, etc. Likewise, of those scoring between 200 and 249, it can be predicted that 87.5% will not become delinquent. Since 80% of predictions will be correct without the index if all cases are called nondelinquent, one would not predict nondelinquency with the index in score intervals over 249. Likewise, it would be best not to predict delinquency for individuals in the intervals under 250 because 20% of predictions will be correct if the base rate is used.

---

[3] The Gluecks' Tables XX-2, 3, 4, 5, (7, pp. 261–262) and their interpretations therefrom are apt to be misleading because of their exclusive consideration of approximately equal base rates of delinquency and nondelinquency. Reiss (13), in his review of the Gluecks' study, has also discussed their use of an unrepresentative rate of delinquency.

It should be emphasized that there are different ways of quantifying one's clinical errors, and they will, of course, not all give the same evaluation when applied in a given setting. "Per cent valid positives" ($=p_1$) is rarely if ever meaningful without the correlated "per cent false positives" ($=p_2$), and clinicians are accustomed to the idea that we pay for an increase in the first by an increase in the second, whenever the increase is achieved not by an improvement in the test's intrinsic validity but by a shifting of the cutting score. But the two quantities p_1 and p_2 do not define our over-all hit frequency, which depends also upon the base rates P and Q. The three quantities p_1, p_2, and P do, however, contain all the information needed to evaluate the test with respect to any given sign or cutting score that yields these values. Although p_1, p_2, and P contain the relevant information, other forms of it may be of greater importance. No two of these numbers, for example, answer the obvious question most commonly asked (or vaguely implied) by psychiatrists when an inference is made from a sign, viz., "How sure can you be on the basis of that sign?" The answer to this eminently practical query involves a probability different from any of the above, namely, the *inverse* probability given by Bayes' formula:

$$H_P = \frac{Pp_1}{Pp_1 + Qp_2}.$$

Even a small improvement in the hit frequency to $H'_T = Pp_1 + Qq_2$ over the $H_T = P$ attainable without the test may be adjudged as worth while when the increment ΔH_T is multiplied by the N examined in the course of one year and is thus seen to involve a dozen lives or a dozen curable schizophrenics. On the other hand, the simple fact that an actual *shrinkage* in total hit rate may occur seems to be unappreciated or tacitly ignored by a good deal of clinical practice. One must keep constantly in mind that numerous diagnostic, prognostic, and dynamic statements can be made about almost all neurotic patients (e.g., "depressed," "inadequate ability to relate," "sexual difficulties") or about very few patients (e.g., "dangerous," "will act out in therapy," "suicidal," "will blow up into a schizophrenia"). A psychologist who uses a test sign that even cross validates at $p_1 = q_2 = 80\%$ to determine whether "depression" is present or absent, working in a clinical population where practically everyone is fairly depressed except a few psychopaths and old-fashioned hysterics, is kidding himself, the psychiatrist, and whoever foots the bill.

"SUCCESSIVE-HURDLES" APPROACH

Tests having low efficiency, or having moderate efficiency but applied to populations having very unbalanced base rates ($P \ll Q$) are sometimes defended by adopting a "crude initial screening" frame of reference, and arguing that certain other procedures (whether tests or not) can be applied to the subset identified by the screener ("successive hurdles"). There is no question that in some circumstances (e.g., military induction, or industrial selection with a large labor market) this is a thoroughly defensible position. However, as a general rule one should examine this type of justification critically, with the preceding considerations in mind. Suppose we have a test which distinguishes brain-tumor from non–brain-tumor patients with 75% accuracy and no differential bias ($p_1 = q_2 = .75$). Under such circumstances the test hit rate H_T is .75 regardless of the base rate. If we use the test in making our judgments, we are correct in our diagnoses 75 times in 100. But suppose only one patient in 10 actually has a brain tumor, we will drop our over-all "success" from 90% (attainable by diagnosing "No tumor" in all cases) to 75%. We do, however, identify 3 out of 4 of the real brain tumors, and in such a case it seems worth the price. The "price" has two aspects to it: We take time to give the test, and, having given it, we call many "tumorous" who are not. Thus, suppose that in the course of a year we see 1000 patients. Of these, 900 are non-tumor, and we erroneously call 225 of these "tumor." To pick up (100) (.75) = 75 of the tumors, *all* 100 of whom would have been called tumor-free using the base rates alone, we are willing to mislabel 3 times this many as tumorous who are actually not. Putting it another way, whenever we say "tumor" on the basis of the test, the chances are 3 to 1 that we are mistaken. When we "rule out" tumor by the test, we are correct 96% of the time, an improvement of only 6% in the confidence attachable to a negative finding over the confidence yielded by the base rates.[4]

Now, picking up the successive-hurdles argument, suppose a major decision (e.g., exploratory surgery) is allowed to rest upon a second test which is infallible but for practically insuperable reasons of staff, time, etc., cannot be routinely given. We administer Test 2 only to "positives" on (screening) Test 1. By this tactic we eliminate all 225 false positives left by Test 1, and we verify the 75 valid positives screened in by Test 1. The 25 tumors that slipped through as false negatives on Test 1 are, of course, not picked up by Test 2 either, because it is not applied to them. Our total hit frequency is now 97.5%, since the only cases ultimately misclassified out of our 1000 seen are these 25 tumors which escaped through the initial sieve Test 1. We are still running only 7½% above the base rate. We have had to give our short-and-easy test to 1000 individuals and our cumbersome, expensive test to 300 individuals, 225 of whom turn out to be free of tumor. But we have located 75 patients with tumor who would not otherwise have been found.

Such examples suggest that, except in "life-or-death" matters, the successive-screenings argument merely tends to soften the blow of Bayes' Rule in cases where the base rates are very far from symmetry. Also, if Test 2 is not assumed to be infallible but only highly effective, say 90% accurate both ways, results start looking unimpressive again. Our net false positive rate rises from zero to 22 cases miscalled "tumor," and we inoperate 67 of the actual tumors instead of 75. The total hit frequency drops to 94.5%, only 4½% above that yielded by a blind guessing of the modal class.

THE SELECTION-RATIO

Straightforward application of the preceding principles presupposes that the clinical decision maker is free to adopt a policy solely on the basis of maximizing hit frequency. Sometimes there are external constraints such as staff time, administrative policy, or social obligation which further complicate matters. It may then be impossible to make all decisions in accordance with the base rates, and the task given to the test is that of selecting a subset of cases which are decided in the direction opposite to the base rates but will still contain fewer erroneous decisions than would ever be yielded by opposing the base rates without the test. If 80% of patients referred to a Mental Hygiene Clinic are recoverable with intensive psychotherapy, we would do better to treat everybody than to utilize a test yielding 75% correct predictions. But suppose that available staff time is limited so that we *can* treat only half the referrals. The Bayes-type injunction to "follow the base rates when they are better than the test" becomes pragmatically meaningless, for it directs us to make decisions which we cannot implement. The imposition of an *externally* imposed selection ratio, not determined on the basis of any maximizing or minimizing policy but by nonstatistical considerations, renders the test worth while.

---

[4] Improvements are expressed throughout this article as *absolute* increments in percentage of hits, because: (*a*) This avoids the complete arbitrariness involved in choosing between original hit rate and miss rate as starting denominator; and (*b*) for the clinician, the person is the most meaningful unit of gain, rather than a proportion *of* a proportion (especially when the reference proportion is very small).

Prior to imposition of any arbitrary selection ratio, the fourfold table for 100 referrals might be as shown in Table 8. If the aim were simply to minimize total errors, we would predict "good" for each case and be right 80 times in 100. Using the test, we would be right only 75 times in 100. But suppose a selection ratio of .5 is externally imposed. We are then forced to predict "poor" for half the cases, even though this "prediction" is, in any given case, likely to be wrong. (More precisely, we handle this subset *as if* we predicted "poor," by refusing to treat.) So we now select our 50 to-be-treated cases from among those 65 who fall in the "test-good" array, having a frequency of $60/65 = 92.3\%$ hits among those selected. This is better than the 80% we could expect (among those selected) by choosing half the total referrals at random. Of course we pay for this, by making many "false negative" decisions; but these are necessitated, whether we use the test or not, by the fact that the selection ratio was determined without regard for hit maximization but by external considerations. Without the test, our false negative rate q_1 is 50% (i.e., 40 of the 80 "good" cases will be called "poor"); the test reduces the false negative rate to 42.5% ($=34/80$), since 15 cases from above the cutting line must be selected at random for inclusion in the not-to-be-treated group below the cutting line [i.e., $20+(60/65)15 = 34$]. Stated in terms of correct decisions, without the test 40 out of 50 selected for therapy will have a good therapeutic outcome; with the test, 46 in 50 will be successes.

TABLE 8

ACTUAL AND TEST-PREDICTED THERAPEUTIC OUTCOME

Test Prediction	Therapeutic Outcome		
	Good	Poor	Total
Good	60	5	65
Poor	20	15	35
Total	80	20	100

Reports of studies in which formulas are developed from psychometrics for the prediction of patients' continuance in psychotherapy have neglected to consider the relationship of the selection ratio to the specific population to which the prediction formula is to be applied. In each study the population has consisted of individuals who were *accepted for* *therapy* by the usual methods employed at an outpatient clinic, and the prediction formula has been evaluated *only* for such patients. It is implied by these studies that the formula would have the same efficiency if it were used for the *selection* of "continuers" from all those *applying* for therapy. Unless the formula is tested on a random sample of applicants who are allowed to enter therapy without regard to their test scores, its efficiency for selection purposes is unknown. The reported efficiency of the prediction formula in the above studies pertains only to its use in a population of patients who have already been selected for therapy. There is little likelihood that the formula can be used in any practical way for further selection of patients unless the clinic's therapists are carrying a far greater load than they plan to carry in the future.

The use of the term "selection" (as contrasted with "prediction" or "placement") ought not to blind us to the important differences between industrial selection and its clinical analogue. The incidence of false negatives—of potential employees screened out by the test who would actually have made good on the job if hired—is of little concern to management except as it costs money to give tests. Hence the industrial psychologist may choose to express his aim in terms of minimizing the false positives, i.e., of seeing to it that the job success *among those hired* is as large a rate as possible. When we make a clinical decision to treat or not to treat, we are withholding something from people who have a claim upon us in a sense that is much stronger than the "right to work" gives a job applicant any claim upon a particular company. So, even though we speak of a "selection ratio" in clinical work, it must be remembered that those cases *not selected* are patients about whom a certain kind of important negative decision is being made.

For any *given* selection ratio, maximizing total hits is always equivalent to maximizing the hit rate for either type of decision (or minimizing the errors of either, or both, kinds), since cases shifted from one cell of the table have to be exactly compensated for. If m "good" cases that were correctly classified by one decision method are incorrectly classified by another, maintenance of the selection ratio entails that m cases correctly called "poor" are also miscalled "good" by the new method. Hence an externally imposed selection ratio eliminates the often troublesome value questions about the relative seriousness of the two kinds of errors, since they are unavoidably increased or decreased at exactly the same rate.

If the test yields a score or a continuously varying index of some kind, the values of p_1 and p_2 are not fixed, as they may be with "patterns" or "signs." Changes in the selection ratio, R, will then suggest shifting the cutting scores or regions on the basis of the relations obtaining among R, P, and the p_1, p_2 combinations yielded by various cuts. It is worth special comment that, in the case of continuous distributions, the optimum procedure is *not* always to move the cut until the total area truncated $= NR$, selecting all above that cut and rejecting all those below. Whether this "obvious" rule is wise or not depends upon the distribution characteristics. We have found it easy to construct pairs of distributions such that the test is "discriminating" throughout, in the sense that the associated cumulative frequencies q_1 and q_2 maintain the same direction of their inequality everywhere in the range

$$\left(\text{i.e., } \frac{1}{N_2}\int_{-\infty}^{x_i} f_2(x)dx > \frac{1}{N_1}\int_{-\infty}^{x_i} f_1(x)dx \text{ for all } x_i\right);$$

yet in which the hit frequency given by a single cut at R is inferior to that given by first selecting with a cut which yields $N_c < NR$, and then picking up the remaining $(NR - N_c)$ cases at random below the cut. Other more complex situations may arise in which different types of decisions should be made in different regions, actually reversing the policy as we move along the test continuum. Such numerical examples as we have constructed utilize continuous, unimodal distributions, and involve differences in variability, skewness, and kurtosis not greater than those which arise fairly often in clinical practice. Of course the utilization of any very complicated pattern of regions requires more stable distribution frequencies than are obtainable from the sample sizes ordinarily available to clinicians.

It is instructive to contemplate some of the moral and administrative issues involved in the practical application of the preceding ideas. It is our impression that a good deal of clinical research is of the "So—what?" variety, not because of defects in experimental design such as inadequate cross validation but because it is hard to see just what are the useful changes in decision making which could reasonably be expected to follow. Suppose, for example, it is shown that "duration of psychotherapy" is 70% predictable from a certain test. Are we prepared to propose that those patients whose test scores fall in a certain range should not receive treatment? If not, then is it of any real advantage therapeutically to "keep in mind" that the patient has 7 out of 10 chances of staying longer than 15 hours, and 3 out of 10 chances of staying less than that? We are not trying to poke fun at research, since presumably almost any lawful relationship stands a chance of being valuable to our total scientific comprehension some day. But many clinical papers are ostensibly inspired by practical aims, and can be given theoretical interpretation or fitted into any larger framework only with great difficulty if at all. It seems appropriate to urge that such "practical"-oriented investigations should be really *practical*, enabling us to see how our clinical decisions could rationally be modified in the light of the findings. It is doubtful how much of current work could be justified in these terms.

Regardless of whether the test validity is capable of improving on the base rates, there are some prediction problems which have practical import only because of limitations in personnel. What other justification is there for the great emphasis in clinical research on "prognosis," "treatability," or "stayability"? The very formulation of the predictive task as "maximizing the number of hits" already presupposes that we intend *not* to treat some cases; since if we treat all comers, the ascertainment of a bad prognosis score has no practical effect other than to discourage the therapist (and thus hinder therapy?). If intensive psychotherapy could be offered to all veterans who are willing to accept referral to a VA Mental Hygiene Clinic, would it be licit to refuse those who had the poorest outlook? Presumably not. It is interesting to contrast the emphasis on prognosis in clinical psychology with that in, say, cancer surgery, where the treatment *of choice* may still have a very low probability of "success," but is nevertheless carried out on the basis of that low probability. Nor does this attitude seem unreasonable, since no patient would refuse the best available treatment on the ground that even it was only 10% effective. Suppose a therapist, in the course of earning his living, spends 200 hours a year on nonimprovers by following a decision policy that also results in his unexpected success with one 30-year-old "poor bet." If this client thereby gains 16×365×40 =233,600 hours averaging 50% less anxiety during the rest of his natural life, it was presumably worth the price.

These considerations suggest that, with the expansion of professional facilities in the behavior field, the prediction problem will be less like that of industrial *selection* and more like that of *placement*. "To treat or not to treat" or "How treatable" or "How long to treat" would be replaced by "What *kind* of treatment?" But as soon as the problem is formulated in this way, the external selection ratio is usually no longer imposed. Only if we are deciding between such alternatives as classical analysis and, say, 50-hour interpretative therapy would such personnel limitations as can be expected in future years impose an arbitrary *R*. But if the decision is between such alternatives as short-term interpretative therapy, Rogerian therapy, Thorne's directive therapy, hypnotic retraining, and the method of tasks (10, 16, 19), we could "follow the base rates" by treating every patient with the method known to have the highest success frequency among patients "similar" to him. The criteria of similarity (class membership) will presumably be multiple, both phenotypic and genotypic, and will have been chosen because of their empirically demonstrated prognostic relevance rather than by guesswork, as is current practice. Such an idealized situation also presupposes that the selection and training of psychotherapists will have become socially realistic so that therapeutic personnel skilled in the various methods will be available in some reasonable proportion to the incidence with which each method is the treatment of choice.

How close are we to the upper limit of the predictive validity of personality tests, such as was reached remarkably early in the development of academic aptitude tests? If the now-familiar $\frac{2}{3}$ to $\frac{3}{4}$ proportions of hits against even-split criterion dichotomies are already approaching that upper limit, we may well discover that for many decision problems the search for tests that will significantly better the base rates is a rather unrewarding enterprise. When the criterion is a more circumscribed trait or symptom ("depressed," "affiliative," "sadistic," and the like), the difficulty of improving upon the base rates is combined with the doubtfulness about how valuable it is to have such information with 75% confidence anyhow. But this involves larger issues beyond the scope of the present paper.

AVAILABILITY OF INFORMATION ON BASE RATES

The obvious difficulty we face in practical utilization of the preceding formulas arises from the fact that actual quantitative knowledge of the base rates is usually lacking. But this difficulty must not lead to a dismissal of our considerations as clinically irrelevant. In the case of many clinical decisions, chiefly those involving such phenotypic criteria as overt symptoms, formal diagnosis, subsequent hospitalization, persistence in therapy, vocational or marital adjustment, and the numerous "surface" personality traits which clinicians try to assess, *the chief reason for our ignorance of the base rates is nothing more subtle than our failure to compute them.* The file data available in most installations having a fairly stable source of clientele would yield values sufficiently accurate to permit minimum and maximum estimates which might be sufficient to decide for or against use of a proposed sign. It is our opinion that this rather mundane taxonomic task is of much greater importance than has been realized, and we hope that the present paper will impel workers to more systematic efforts along these lines.

Even in the case of more subtle, complex, and genotypic inferences, the situation is far from hopeless. Take the case of some such dynamic attribution as "strong latent dependency, which will be anxiety-arousing as therapy proceeds." If this is so difficult to discern *even during intensive therapy* that a therapist's rating on it has too little reliability for use as a criterion, it is hard to see just what is the value of guessing it from psychometrics. If a skilled therapist cannot discriminate the personality characteristic after considerable contact with the patient, it is at least debatable whether the characteristic makes any practical difference. On the other hand, if it can be reliably judged by therapists, the determination of approximate base rates again involves nothing more complex than systematic recording of these judgments and subsequent tabulation. Finally, "clinical experience" and "common sense" must be invoked when there is nothing better to be had. Surely if the q_1/q_2 ratio for a test sign claiming validity for "difficulty in accepting inner drives" shows from the formula that the base rate must not exceed .65 to justify use of the sign, we can be fairly confident in discarding it for use with *any* psychiatric population! Such a "backward" use of the formula to obtain a maximum useful value of P, in conjunction with the most tolerant common-sense estimates of P from daily experience, will often suffice to answer the question. If one is really in complete ignorance of the limits within which P lies, then obviously no rational judgment as to the probable efficiency of the sign can be made.

ESTIMATION VERSUS SIGNIFICANCE

A further implication of the foregoing thinking is that the exactness of certain small sample statistics, or the relative freedom of certain nonparametric methods from distribution assumptions, has to be stated with care lest it mislead clinicians into an unjustified confidence. When an investigator concludes that a sign, item, cutting score, or pattern has "validity" on the basis of small sample methods, he has rendered a certain very broad null hypothesis unplausible. To decide, however, whether this "validity" warrants clinicians in using the test is (as every statistician would insist) a further and more complex question. To answer this question, we require more than knowledge that $p_1 \neq p_2$. We need in addition to know, with respect to each decision for which the sign is being proposed, whether the appropriate inequality involving p_1 p_2, and P is fulfilled. More than this since we will usually be extrapolatin to a somewhat different clinica population, we need to know whethe altered base rates P' and Q' wil falsify these inequalities. To do thi demands *estimates* of the test parame ters p_1 and p_2, the setting up of confidence belts for their difference $p_1 - p_2$ rather than the mere proof of their nonidentity. Finally, if the sign is a cutting score, we will want to consider shifting it so as to *maintain* optimal hit frequency with new base rates. The effect upon p_1 and p_2 of a contemplated movement of a critical score or band requires a knowledge of distribution form such as only a large sample can give.

As is true in all practical applications of statistical inference, nonmathematical considerations enter into the use of the numerical patterns that exist among P, p_1, p_2, and R. But "pragmatic" judgments initially require a separation of the several probabilities involved, some of which may be much more important than others in terms of the human values associated with them. In some settings, over-all hit rate is all that we care about. In others, a redistribution of the hits and misses even without much total improvement may concern us. In still others, the proportions p_1 and q_2 are of primary interest; and, finally, in some instances the confrontation of a certain increment in the absolute frequency (NPp_1) of one group identified will outweigh all other considerations.

Lest our conclusions seem unduly pessimistic, what constructive suggestions can we offer? We have already mentioned the following: (*a*) Searching for subpopulations with different base rates; (*b*) successive-hurdles testing; (*c*) the fact that even a very small *percentage* of improvement may be worth achieving in certain crucial decisions; (*d*) the need for systematic collection of base-rate data so that our several equations can be applied. To these we may add two further "constructive" comments. First, test research attention should be largely concentrated upon behaviors having base rates nearer a 50–50 split, since it is for these that it is easiest to improve on a base-rate decision policy by use of a test having moderate validity. There are, after all, a large number of clinically important traits which do not occur "almost always" or "very rarely." Test research might be slanted more toward them; the current popularity of Q-sort approaches should facilitate the growth of such an emphasis, by directing attention to items having a reasonable "spread" in the clinical population. Exceptions to such a research policy will arise, in those rare domains where the pragmatic consequences of the alternative decisions justify focusing attention almost wholly on maximizing Pp_1, with relative neglect of Qp_2. Secondly, we think the injunction "quit wasting time on noncontributory psychometrics" is really constructive. When the clinical psychologist sees the near futility of predicting rare or near-universal events and traits from test validities incapable of improving upon the base rates, his clinical time is freed for more economically defensible activities, such as research which will improve the parameters p_1 and p_2; and for *treating* patients rather than uttering low-confidence prophecies or truisms about them (in this connection see **12**, pp. vii, 7, 127–128). It has not been our intention to be dogmatic about "what is worth finding out, how often." We do suggest that the clinical use of patterns, cutting scores, and signs, or research efforts devoted to the discovery of such, should always be evaluated in the light of the simple algebraic fact discovered in 1763 by Mr. Bayes.

SUMMARY

1. The practical value of a psy-

chometric sign, pattern, or cutting score depends jointly upon its intrinsic validity (in the usual sense of its discriminating power) and the distribution of the criterion variable (base rates) in the clinical population. Almost all contemporary research reporting neglects the base-rate factor and hence makes evaluation of test usefulness difficult or impossible.

2. In some circumstances, notably when the base rates of the criterion classification deviate greatly from a 50 per cent split, use of a test sign having slight or moderate validity will result in an *increase* of erroneous clinical decisions.

3. Even if the test's parameters are precisely known, so that ordinary cross-validation shrinkage is not a problem, application of a sign within a population having these same test parameters but a different base rate may result in a marked change in the proportion of correct decisions. For this reason validation studies should present trustworthy information respecting the criterion distribution in addition to such test parameters as false positive and false negative rates.

4. Establishment of "validity" by exact small sample statistics, since it does not yield accurate information about the test parameters (a problem of estimation rather than significance), does not permit trustworthy judgments as to test usefulness in a new population with different or unknown base rates.

5. Formulas are presented for determining limits upon relations among (*a*) the base rates, (*b*) false negative rate, and (*c*) false positive rate which must obtain if use of the test sign is to improve clinical decision making.

6. If, however, external constraints (e.g., available staff time) render it administratively unfeasible to decide all cases in accordance with the base rates, a test sign may be worth applying even if following the base rates *would* maximize the total correct decisions, were such a policy possible.

7. Trustworthy information as to the base rates of various patient characteristics can readily be obtained by file research, and test development should (other things being equal) be concentrated on those characteristics having base rates nearer .50 rather than close to .00 or 1.00.

8. The basic rationale is that of Bayes' Theorem concerning the calculation of so-called "inverse probability."

REFERENCES

1. American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education, Joint Committee. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, 51, 201–238.
2. Anastasi, Anne, & Foley, J. P. *Differential psychology.* (Rev. Ed.) New York: Macmillan, 1949.
3. Bross, I. D. J. *Design for decision.* New York: Macmillan, 1953.
4. Danielson, J. R., & Clark, J. H. A personality inventory for induction screening. *J. clin. Psychol.*, 1954, 10, 137–143.
5. Dorken, H., & Kral, A. The psychological differentiation of organic brain lesions and their localization by means of the Rorschach test. *Amer. J. Psychiat.*, 1952, **108**, 764–770.
6. Duncan, O. D., Ohlin, L. E., Reiss, A. J., & Stanton, H. R. Formal devices for making selection decisions. *Amer. J. Sociol.*, 1953, **58**, 573–584.
7. Glueck, S., & Glueck, Eleanor. *Unraveling juvenile delinquency.* Cambridge, Mass.: Harvard Univer. Press, 1950.
8. Goodman, L. A. The use and validity of a prediction instrument. I. A reformulation of the use of a prediction instru-

ment. *Amer. J. Sociol.*, 1953, **58**, 503–509.

9. Hanvik, L. J. Some psychological dimensions of low back pain. Unpublished doctor's thesis, Univer. of Minnesota, 1949.
10. Herzberg, A. *Active psychotherapy.* New York: Grune & Stratton, 1945.
11. Horst, P. (Ed.) The prediction of personal adjustment. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48, 1–156.
12. Meehl, P. E. *Clinical versus statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954.
13. Reiss, A. J. Unraveling juvenile delinquency. II. An appraisal of the research methods. *Amer. J. Sociol.*, 1951, **57**, 115–120.
14. Rosen, A. Detection of suicidal patients: an example of some limitations in the prediction of infrequent events. *J. consult. Psychol.*, 1954, **18**, 397–403.
15. Rotter, J. B., Rafferty, J. E., & Lotsof, A. B. The validity of the Rotter Incomplete Sentences Blank: high school form. *J. consult. Psychol.*, 1954, **18**, 105–111.
16. Salter, A. *Conditioned reflex therapy.* New York: Creative Age Press, 1950.
17. Taulbee, E. S., & Sisson, B. D. Rorschach pattern analysis in schizophrenia: a cross-validation study. *J. clin. Psychol.*, 1954, **10**, 80–82.
18. Thiesen, J. W. A pattern analysis of structural characteristics of the Rorschach test in schizophrenia. *J. consult. Psychol.*, 1952, **16**, 365–370.
19. Wolpe, J. Objective psychotherapy of the neuroses. *S. African Med. J.*, 1952, **26**, 825–829.

Received for early publication December 24, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 3, 1955

INTRA-INDIVIDUAL RESPONSE VARIABILITY[1]

DONALD W. FISKE AND LAURA RICE[2,3]
University of Chicago

THE PROBLEM

This review of the literature is a foundation stone for a research program on intra-individual variability, the variability of an individual's behavior from one time to another. The program's objectives include the exploration of the phenomena of variability within an individual's behavior, with consequent implications for theory and practice in psychometrics and in personality. The program seeks light on several questions. Can we partial out from the conventional error variance of psychometrics a component of variance over time which is associated with the individual? (Such a component would probably make different proportional contributions to the variance of scores for different individuals.) Are there variability factors analogous to the well-known factors of level scores in mental abilities, interests, and personality? How can the concept of intra-individual variability contribute to problems of the prediction of behavior? What is the significance of this concept for the study of personality and for personality theory? Is variability within a class of behavior associated with degree of integration within an area of personality? Are the roots of variability in the neural or physiological functioning of the organism?

The problem of intra-individual variability has not been subjected to systematic conceptualization. Hence it is difficult to set up definite criteria for inclusion in this review and to organize the numerous but highly disparate studies which seem pertinent.

As a means of structuring the problem, we shall first delineate a model instance. *Pure intra-individual variability is defined as the difference between the two responses of an individual at two points in time under the following conditions:* (a) *the individual is exposed each time to the same stimulus or to objectively indistinguishable stimuli;* (b) *the total situation in which the responses are made is the same on both occasions.* It is doubtful whether such an abstract case ever exists. Guthrie (89) argues that the exact situation is never reproduced. Several problems await answers. What degree of homogeneity between stimuli and between situations is required before we can assume that psychological equivalence exists? In judging equivalence, what are the boundaries (spatial, temporal, and psychological) of a "total" situation?

In this formulation, the meanings of two words need clarification. We are using "situation" to refer to the total immediate environment of the organism. Strictly speaking, it should include the stimulus on which we focus our attention. The situation can be defined as embracing all factors external to the organism, which affect responses.

---

[1] This research was supported in part by the United States Air Force under Contract Number AF 18(600)-601, monitored by the Director of Research, 3300th Research and Development Group (Personnel Research Laboratory), Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] We are indebted to Dr. Murray Glanzer for his helpful comments on a draft of this paper, and to our colleagues: Mr. Fred Zimring helped search the literature; Mr. Jerry Osterweil and Mr. Shib Mitra contributed pertinent criticisms.

[3] This review includes publications through 1953.

At this point, "response" is not given a formal definition. When we examine the difference between two responses, we may restrict ourselves to any one of several attributes, such as magnitude, intensity, latency, or quality.

This paradigm involves the further assumption that *the order of the two responses is immaterial.* This requirement implies that the *responses show no systematic trend over time,* due to such processes as learning, fatigue, etc., i.e., that the response is not a function of time. This assumption also implies that the *second response is not affected by either the first response or the first presentation of the stimulus.* We may call this pure case spontaneous variability, as opposed to reactive or adaptive variability.

Finally, there is an assumption which underlies this entire research program: *intra-individual response variability is not random;* it is a lawful phenomenon. The variability of one individual's responses to one stimulus is determined by more or less enduring factors within the individual. Two postulates can be derived from this assumption. (*a*) Given the same stimulus in the same situation, the difference between an individual's responses at two points in time is related to the difference between his two responses (to that stimulus) at two other points in time. (*b*) Furthermore, the difference between his responses to one stimulus is related to the difference between responses to at least one other stimulus, objectively distinguishable from the first one. Presumably, the magnitude of this latter relationship is a function of the similarity of the stimuli. This general fundamental assumption is obviously testable. On the basis of our preliminary studies and of research cited below, we consider that this assumption has been verified.

Data that exactly fit this paradigm are rare. Usually we shall be concerned with variability summed over several stimuli (e.g., number of test responses changed on retest) or with variability in a composite score based on the sum of several responses (e.g., change in total score on an instrument measuring intelligence, performance, interest, etc.).

In the rest of this paper, we shall be primarily concerned with the pure case or approximations to it. Such instances will be called *Type I* or spontaneous variability.

Second in importance to us is *Type II*, the case where all conditions and assumptions for the pure case, Type I, are met with the exception that the sequence of responses shows some pattern or order, other than a monotonic function of time. The simplest example is the alternation of responses. In this class fall instances where the second response is affected by the first response or the first presentation of the stimulus (as in alternation) and also instances where the differences between successive responses are less marked—e.g., where cycles or oscillations are present. We shall not attempt a comprehensive coverage of the literature on Type II variability, especially of the latter kind.

Most instances of Type II variability can be classed as reactive variability: the change in response is determined in part by the organism's reaction to the stimulus it has recently reacted to and/or its reaction to its preceding response.

We do not know yet whether Type I and Type II are actually different in practice. Nevertheless, the concepts underlying these classes are quite distinct: Type II variability is composed of variability from the same sources as Type I, but includes in addition reactive or periodic variability. Thus all the determinants of Type I variability are present in Type II, but Type II has one extra and major determinant associated with temporal order. For the present, we are making the distinction on an empirical basis: Is there evidence that the responses show some simple ordering? It is possible that the assumption of no systematic trend over time which we make for Type I is unjustified. Cases exist, however, where the order of the responses can be treated as having a negligible effect.

A third class of problems, *Type III*, differs from the pure Type I case in that objectively different stimuli are presented on the two occasions or the background situation is changed. Here, the focus is usually on the appropriateness of the change in response. Is the difference between the subject's two responses too small—does he fail to "adapt" to the change of the stimulus (as in some studies of rigidity)? Is the difference too large—for example, does he over-react to stress?

It is obvious that the Type III case can be taken to include any comparison between responses in two situations which is not included in either Type I or Type II. Hence, our discussion of Type III variability will mention only studies which may help us to understand Type I phenomena. We shall not consider "scatter" or profile variability on different aptitudes. In general, in this paper we are attempting comprehensive coverage of only Type I variability. Our explorations into related topics are solely to clarify our central problem and to obtain leads for attacking it.

As far as we know, there have been no recent reviews of the literature on Type I variability. Some of the earlier studies are examined by Allport and Vernon (**3**, pp. 124–128). Solomon (**201**) discusses research on many topics related to variability, especially work on the avoidance of repetition of response (Type II variability). Glanzer (**77**) examines much of the literature on alternation. An extensive list of papers on behavior in guessing is provided by Senders and Sowards (**194**). Other surveys of the literature on related topics are mentioned in subsequent sections of this paper.

The sections on these three types of variability are followed by a survey of studies of the correlates of variability. Here are data from which may be gleaned some preliminary notions about the origin and structure of variability.

In this paper, unless otherwise qualified, the term "variability" always refers to intra-individual response variability. Any references to conventional interindividual or group variability (i.e., "individual differences") will be explicitly indicated.

VARIABILITY OF ORGANIC PROCESSES

It is obvious that spontaneous variability (Type I) is produced by factors within the organism. In the pure case, the stimulus and the external situation are unchanged, leaving only the organism as the locus for determinants of variability. Presumably there must be some variation or change in organic processes underlying response variability. Each of these processes may itself be variable in its functioning, or it may be regular over time—we need make no assumption as to the character of this organic variation. However, it would be poor strategy to embark on a study of response variability without evidence that there exists some variability in the functioning of organic processes, independent of direct effects of changes in the external environment. This section will cite work which supports such a belief.

Loeb (**140**, p. 622) points out that interindividual differences and the nonpredictability of individual responses increase as the structure of the organism, especially the nervous system, becomes more complex. Portraying the organism as an active system, Bertalanffy argues that movements (or responses) are influenced by "spontaneous fluctuations in the excitations of nerve centers" (**10**, p. 167); therefore, from his point of view, reactions are determined by the changing internal situation within the organism, not directly by the stimulus outside. Rashevsky elaborates on these spontaneous fluctuations: " . . . in general we must expect spontaneous fluctuations of excitation to occur in the central nervous system. Such fluctuations may be due to fluctuations of metabolic activity or to excitation carried to a given region from a number of other regions of the brain, which are randomly excited by the stream of incoming exteroceptive as well as proprioceptive and enteroceptive stimuli" (**177**, p. 166). (Cf. **130**.) Chocholle (**32**) discusses causes of fluctuation in auditory reaction time which are of central origin, such as the state of the sensory areas.

The spontaneous activity of the nervous system presumably lies behind the extraordinary phenomena reported by Heron, Bexton, and Hebb (**105**). Subjects kept for a day or more in a condition involving a marked reduction in sensory stimulation reported "visual imagery, dreamlike in vividness," akin to hallucination.

Differences in intra-individual variability in biochemistry are emphasized by Jellinek (**123**). Persky (**170**) has compared the average intra-individual variabilities for three biochemical stress indices. A relationship between emotional instability and variability in blood constituents has been reported by Hammett (**100**) and by Goldstein (**82**).

In analyzing variation, Crozier and Hoagland (**42**) distinguish between errors of observation and recording and the real variation found in the responses of living organisms. The latter can be considered to be "unpredictableness" due to the extreme complexity of organic systems. They cite experimental evidence that variability is reduced by an increase in excitation and motor output. We may entertain the tentative hypothesis that, up to some point, as the demands of a stimulus situation increase the mobilization of the organism, the variability of response decreases.

Empirical data on neural and physiological variability are available. Blair and Erlanger (**15**) found that neural response thresholds and reaction latencies of individual axon fibers vary spontaneously from instant to instant. Herrington (**106**) measured basal metabolic rate, systolic blood pressure, respiration rate, pulse rate, and rated activity 45 times over 90 days for 11 subjects. With the exception of intercorrelations involving BMR, the median intercorrelation of the standard deviations was as high as the median intercorrelation between means. Furthermore, the ratings on average activity level were related to both the mean levels and the variability of the three physiological measures (again excluding BMR). From an extensive study of basal metabolic rates, Harmon (**101**) showed that BMR measurements taken under standard conditions indicate considerable variation from day to day.

In his studies of biological intelligence, Halstead (**95**) has noted that the average deviations of normal subjects on critical fusion frequency show cyclical variation over time. Schmidtke (**190**) observed that performance on tests of *cff* varied with time of day, the amount of vacillation showing individual differences. McNemar studied variability in critical fusion frequency on different days and under different conditions. She concluded that "individuals do not exhibit the same degree of variability in response errors under all conditions of measurement" (**144**, p. 21). However, many of the intercorrelations between variability measures from different test conditions on the same day were significant at the .05 level. Variability for the same condition on different days was found to be relatively unstable. This study illustrates a major problem in measuring variability. Were the six observations per trial sufficient to yield a stable value? If not, would it be possible to increase the number of observations and still eliminate any effects of fatigue and lowered motivation?

Fluctuations of minimal sensory stimulations at threshold have been studied for many years. Early work in this area is summarized by Guilford (**87**), who interpreted them as fluctuations of attention, but Bills (**13**) objected to this interpretation because the periodicities were different for different sense organs. (Cf. **46**, discussed below.)

In this review, we have not attempted systematic coverage of the literature on fluctuations of attention, partly because much of it ignores both inter- and intra-individual differences. For example, Butorin (**22**) studied variation in speed of addition as a measure of stability of attention.

Brunswik argues that the organism must be flexible because it has such incomplete knowledge of its environment: "Ambiguity of cues and means relative to the vitally relevant objects and results must find its counterpart in an ambiguity and flexibility of the proximal-peripheral mediating processes in the organism" (**21**, pp. 257–258). The theme that the capacity to vary responses is essential to individual development and survival is implicit or explicit in many publications. Not only must the individual respond differently to different situations but he must also vary his response to the same situation in order to adapt, i.e., to improve his adjustment to the situation.

PSYCHOMETRIC ASPECTS OF VARIABILITY

Before proceeding with the main body of this paper, the review of studies of individual variability in response, we should consider briefly the various possible measures of variability and some problems associated with this measurement. This discussion is necessary because the topic has not been systematically studied before.

Approaches to the Measurement of Variability

The phenomena of variability are usually viewed negatively. Efforts are usually made to minimize the extent of intra-individual variability. Thus one ordinarily seeks high test-retest reliability (i.e., low variability) and eliminates stimulus items to which inconsistent (varying) responses are made. Our problem can be viewed as the measurement of the unreliability of the individual. Guttman (**91**) points out that there are three sources of psychometric variation: persons, items, and trials. He formulates an expression for the variance of an individual on an item. In his equation for reliability he includes an error variance term which is the mean of the error variances for individuals. Thus he does not assume that individuals are equally unreliable, but rather allows for individual differences in unreliability. Guttman (**92**, **93**) has also developed formulas for the reliability of qualitative (categorical) data.

In discussing reliability, Coombs (**37**) also considers the individual reliability of the same item over time. This formulation is developed in a more generalized form in his *Theory of Psychological Scaling* (**38**). An earlier paper (**35**) considers the problem of obtaining a dispersion score for an individual. Both Coombs and Guttman, however, give formal recognition to the concept without solving the practical problems involved in its measurement.

In psychometric theory, the assumption is sometimes made that errors of measurement for two tests or for the same test on two occasions are uncorrelated. This assumption has been questioned by several writers. Thouless (**214**) discusses and demonstrates the fluctuation of a mental function over time. The variation over time of a person's scores around his true score is noted by Ferguson (**68**) and by Brown and Thomson (**18**).

Cattell (**24**, p. 105) holds that dynamic traits fluctuate even more widely from day to day. (Cf. also **25** and **29**.)

An intensive study of components of test unreliability has been made by Fagin (**65**). He demonstrates that this unreliability is composed of random error variance plus quotidian variance (cf. **228**, below) or consistent personal variation. For trait items, more quotidian variance was found in more reliable items, but this relationship was not present in interest items. He concludes that the proportion of quotidian variance can be estimated and that quotidian scores for people are reliable.

A series of papers by Glaser (**78**, **79**, **80**, **81**) analyzes changes in the responses on retest. Defining inconsistency as the number of responses changed on retest, he finds that the three intercorrelations of inconsistency scores on three tests (intelligence, interests, personality) are uniformly low although two are statistically significant (**78**). He indicates that the number of changes between one pair of two trials (out of three) is related to the number between another pair and that the mean difficulty level of items with inconsistent (changed) responses correlates highly with performance (**80**, **81**). (Cf. Yoshioka's finding [**231**] that rats varied their choice more when the discrimination was more difficult. In a situation where two paths of different length were available, the proportion of choices of the shorter ["correct"] path was a function of the ratio of their lengths, not their absolute lengths.) Glaser shows further that inconsistency scores have no relationship to level scores for a test with an effective range appropriate to the group tested. However, if a test is too easy or too difficult for a group and yields a markedly skewed distribution of scores, a relationship will be found between level score and consistency score. He also reports results consistent with Mosier's finding (**157**) of high split-half reliability for the difficulty value of the median error. (See also **229**, discussed below.)

Cox (**39**) obtained substantial negative correlations between variability from day to day and initial ability on a motor task. The relation between variability and improvement varied from task to task.

The possibility of a relationship between a variability score and the level score on the same responses must always be kept in mind. Where the level score contributes to or is a determinant of the variability score, its effects should usually be partialed out. However, it is conceivable that under some conditions, the variability itself may affect the level score.

McReynolds (**145**) developed a measure of consistency which involved level of difficulty. His subjects were asked whether they could "see" a given concept in an indicated area on a Rorschach card. Ordering the concepts in terms of difficulty, he developed a score based on the relative disorder of the subject's replies, i.e., deviations from saying "yes" to all concepts below a particular level and "no" to all above. This score approximates Coombs' dispersion score of items for one individual (**35**).

Given a series of repeated measurements for the same individual, several measures of their variability are available and have been used: the standard deviation (**42**, **136**); the average deviation and the range (**166**). Measures of profile similarity can frequently be used as measures of variability: e.g., *D*, based on squared differences between paired responses, as discussed by Cronbach and Gleser (**41**). Such a measure may, however, be related to the means and sigmas of the response distributions. Correlation coefficients may also be appropriate measures in some instances.

Noting that variation from day to day occurs even when external conditions are well controlled, Woodrow (**228**) has suggested the concept of "quotidian variability," which may be used to describe individuals and also to check on the stability of internal conditions during an experiment. He recommends a ratio where the numerator is the standard deviation of the daily means and the denominator is the average of the standard deviations for each day divided by the square root of the number of trials per day. This formula measures variation from one day to the next in terms of variation within one occasion. However, we must not overlook the individual differences in intra-occasion variability.

Some investigators (**64**, **72**, **116**) have utilized measures that take into account the relative position of the responses.

Another method of studying sequences of responses is spectral analysis, which yields a profile of the relative contributions to the total variance (or oscillations of performance) of each of the possible component waves. Abelson (**1**) demonstrated its use on a perceptual-motor task and compared it to a conventional measure, the variance. The correlation was .08, indicating the essential independence of these two measures of variability derived from the same set of responses.

On the other hand, De Valois (**50**) found substantial relationships (of the order of .70) among three measures, even when each measure was based on a different set of responses. Using a five-unit maze, his measures were number of different paths used, number of specific choices changed from the choice on last trial, and number of times a third, shorter alternative was used when it was opened up.

The intercorrelation of constancy scores (unchanged responses on retest) has been shown by Dunlap (**58**) to be in the order of .40 for the several sections of his Preference Blank. The constancy indices had small positive relationships with the level scores for corresponding areas and even less relationship with intelligence.

Many general problems of psychological measurement must be reconsidered in developing measures of variability. For example, there is the assumption of additivity of responses which is generally accepted for level scores, in spite of an occasional strong protest. Since we obviously wish to avoid studying the variability of each separate item, we can develop a total change score by counting the number of changed responses (cf. **78**, **81**, **104**, **138**). This measure is similar to Zubin's measure of like-mindedness (**233**).

Such a sum-of-changes score must be distinguished from a score based on change-in-total or level. Thus we can count the number of answers changed on a multiple-choice test or we can compute the change in total score from one trial to the rest (cf. **136** and **206**). But a change-in-total score is an index to change in some posited trait, not the tendency to change specific responses.

Another distinction, which was not made in the last section, should be made explicit. Our basic model deals with change in response when the same stimulus is repeated. But there is also the concept of discrepancy between response to homogeneous stimuli, i.e., to stimuli so objectively similar that the same response is expected (cf. Type III variability). Applicable concepts here are Coombs' homogeneity (**37**) and dispersion score over items (**35**, **36**). Any attempt to measure variability through discrepancies in responses to homogeneous items must be based on the assumption that the degree of homogeneity of two items is constant for all individuals. Otherwise, the discrepancy could be a consequence not of individual variability, but of the heterogeneity of the items for that individual. Thus the study by McReynolds (**145**), mentioned above, assumes that the concepts had the same order of difficulty or plausibility for all subjects. Stated in another way, McReynolds did not directly observe intra-individual variability or consistency but rather the extent to which each subject's order of difficulty (as found on one occasion) corresponded with the order of average difficulty for a group.

The Consistency of Variability (Relationships with Time)

Variability involves the difference between two responses. A measure of variability is usually computed from several such differences between paired responses or from the several deviations from the central tendency of a series of responses. The several responses may be made on one occasion, on two occasions, or on several occasions.

Are measures of variability consistent? Does variability within one occasion show internal consistency? Is variability on one occasion related to variability on another? Are variability measures based on comparisons of responses on two or more occasions internally consistent? Do variability measures show systematic trends over time? While definitive answers to all these questions cannot be given at this time, it is appropriate to examine the available evidence.

Some work on variability suggests that variability has fair consistency over time, i.e., variability measures from different occasions are related to each other. Thorndike (**213**) reported continued variability in the repeated spellings of the same sound in nonsense syllables. In his work on covariation of efficiency of performances on several tasks, Asch (**7**) noted that an individual's variability remained constant as learning progressed. To increase the stability of his measures of variability, Fryer (**73**) discarded the first and the last of ten trials per day. In our analyses of his data, the individual standard deviations for the second, third, and fourth days showed appreciable intercorrelations (.48 to .82) with each other but not with the standard deviations for the first day. Flügel's measure of oscillation (**71**) was highly consistent from the first half to the second half of 46 daily sessions. (Cf. Lovell [**142**] discussed below, and De Valois [**50**] who used an approach analogous to Cronbach's coefficient of stability and equivalence [**40**, pp. 69–70].)

Scores for fluctuation in attitudes and sentiments between two sessions a day apart were related to scores for fluctuations between sessions a month apart (Cattell, **23**): the correlation was .47 for children and .77 for adults. Preliminary studies by the authors suggest that such fluctuation scores are correlated with the tendency to make extreme ratings, but the relationship is not an artifact: unlike those giving many extreme ratings, individuals placing most of their ratings in the center do not change their occasional extreme responses any more than they change their more moderate, center responses.

Cummings (**43**) measured variability within three-minute periods. For three diverse tests, such scores were highly consistent over ten periods. When the ten periods were on the same day, the average scores for the three tests had only low intercorrelations. However, when the ten periods were on ten different days, the intercorrelations between the average variability scores (for each of the three tests) were sufficiently high to suggest a common variability factor.

The reliability of a variability score may also be a function of the nature of the responses studied. Allport and Vernon (**3**) reviewed earlier literature and concluded that variability measures derived from "raw physical scores," such as tapping, are reliable but highly specific, while lower reliability is found for variability measures derived from "more complex tests of intelligence or personality whose initial reliability is not high" (**3**, p. 128). Also, according to them, Reymert (**178**) found "that the measure of variability in reaction time failed to correlate with the variabilities of more integrated activities such as reading and counting" (**3**, p. 132).

A fundamental consideration is the time period between responses. Dudek (**56**) considers that the variability of individuals often contributes to inconsistency in test scores. He finds that the Spearman-Brown prophecy formula holds for the split-half reliability of a test on a single occasion but may not hold for test-retest reliability. Paulsen (**168**, **169**) also found that split-half reliabilities did not predict test-retest values and reported that intertrial correlations decreased as the number of intervening trials increased. (The latter tendency was also noted by Hertzman [**107**] and by McNemar [**144**].) The studies by Dudek and Paulsen used perfectly homogeneous tests of steadiness—the "stimuli" remained the same throughout. Thus, variability over time is related to, but is not identical with, inconsistency of performance on a single occasion.

Variability over time may occur under either of two conditions, defined in terms of rate of organic fluctuation. (*a*) While the organism may respond consistently within any short test session, it may function differently on different occasions clearly separated in time. (*b*) On the other hand, the organic processes determining the response to a given stimulus may be in a continual state of flux, such that two very different responses are made just a few seconds apart. If this condition exists, then the responses within a given occasion are a random sample of the individual's responses to the stimulus and variation over time can be predicted from variation within an occasion. If the first condition holds, variation within an occasion may be unrelated to variation over time.

On the other hand, many studies show that variability itself may be a function of the number of trials: variability may increase or decrease as the series of responses continues. Thorndike (**210**) reports a slight reduction in variability in drawing hundreds of lines of specified length. With reward or punishment, the average error (variability) in a discrimination task was shown by Hamilton (**97**) to decrease as the number of trials increased. Lashley (Crozier and Hoagland, **42**) found that in archery practice, the standard deviation decreased but the relative variability was constant. For addition problems, Flügel's subjects showed increased absolute oscillation but decreased percentage oscillation over 46 daily sessions (**71**). Sarvis (**188**) noted that rhythms in speed of tracing mazes tended to disappear over time. In Hall's study (**94**), rats showed less variability on later trials in a five-alternative maze. Vacillation decreased with training in the experiment of Mowrer and Jones (**159**). Variability during conditioning, extinction, and reconditioning was studied by Antonitis (**5**). He found that variability of response decreased as a function of the number of reinforcements during conditioning, increased rapidly during extinction trials, and decreased during reconditioning below the level for conditioning trials. Greater variability in output on a continuing task is associated with increased fatigue, according to Bills. This represents a "breakdown in the controlling set" (**14**, p. 54).

Repeated self-descriptive *Q* sorts have been analyzed by D. M. Taylor (**206**). In a series of sorts, later ones showed higher intercorrelations than earlier ones. The consistency of self concept over time correlated .33 with the relative positiveness of the initial self-description. Sorts for self were less consistent than sorts for ideals. (Cf. **117**, discussed below.)

The *Studies in Expressive Movement*, by Allport and Vernon (**3**), provide some highly provocative data. They tested 25 subjects on a wide variety of tests in three different sessions. Although they noted that the reliability of level scores was lower between sessions than within sessions, they utilized individual variability scores based on measures from one, two, or three sessions, thus confounding intrasession and intersession variability. Furthermore, while some of the scores were based on variability of response with the stimulus situation essentially unchanged, other variability scores reflected the difference between responses to similar but clearly nonidentical stimuli. Also, they utilized several different types of scores, such as average deviation, coefficient of variability, etc. For these reasons, their data are difficult to interpret.

The average intercorrelation among 11 rank orders on variability was .02. They concluded, however, that there was evidence for general variability and specific variability components. The first correlated .26 with their Emphasis factor and the second correlated .38 with Centrifugality.

On the other hand, the data published by Hunt (**116**) show that for both normals and schizophrenics, the rank orders of individuals on variability over 15 trials five days apart maintain some consistency from test to test. (For each group, our analysis of his data yielded a chi square from the analysis of variance of ranks with a p value less than .10.)

Set toward task. Up to this point, we have assumed that the task and the individual's orientation toward it are constant. Some studies, however, suggest that the individual's set toward the task may change as the task proceeds (**73**, **120**), and that differences in set may affect variability.

Abelson (**1**) reports that the individual variances tended to decrease during the course of the first session but tended to increase during the second session. He interprets the first trend as due to accommodation or learning and the second as due to boredom.

Changes in set occurring later in a series of tests may also affect variability. For rats trained (with food reward) to make brightness discriminations in a maze with a variety of indirect paths, Maier (**146**) noted tendencies for more errors in later trials on the same day and on later days for the same pattern. He attributed the errors to inattention rather then ignorance. Variability increased when the correct path was left unchanged for several days. Taylor (**207**) reported a study by Danziger in his laboratory in which the variability of the autokinetic effect remained unchanged over time for volunteer subjects but decreased markedly for paid subjects. In Philpott's work (**172**), any tendency toward increase in cycle length might well be due to fatigue or boredom (cf. also **127**).

The problem of set in relationship to the variability of schizophrenics will be discussed in the section on "Personality types and diagnostic categories."

Summary

The measurement of variability involves a number of problems. Some of them are similar to those present in conventional psychological measurement which emphasizes level scores. Other problems are peculiar to the study of variability. We must guard against implicitly carrying over the assumptions for level scores as we explore this less familiar area.

Because research on variability has been scattered and unsystematic, a number of different measures have been tried, each implying its own concept of variability. We are not yet in a position to decide upon the ideal measure (if one exists).

Are measures of variability within one occasion consistent or stable over time? The answer appears to be a qualified one: under highly constant conditions, variability measures within different occasions may agree well; however, variability measures are readily affected by the set of the subject toward the task, and therefore by changes in such set. On the other hand, measures of variability *over* occasions may be reasonably stable.

STUDIES OF SPONTANEOUS (APERIODIC) VARIABILITY

In this section we shall consider studies reporting variability which fits closely to our basic paradigm (Type I, spontaneous variability). While such variability has been observed in a wide range of behavioral responses, it has usually been classified as error variance without further interpretation. It is our thesis that this type of variability is a lawful characteristic of an individual in a situation, and that its investigation will enlarge our understanding of behavior.

The need and the capacity for variation have been distinguished. Many years ago, Tolman emphasized the initial exploratory tendencies in the rat (**216**) and creative instability, the capacity to break out into new lines of activity (**217**). Hilgard (**108**) points out that variability may mean a need or preference for variety and also an ability to vary behavior in a given situation. Both meanings refer to variables underlying the observed phenomena. The need for variety involves a reaction to previous stimuli or responses (cf. Type II). In our approach, we assume that the potentiality for changing a response always exists. Maier (**147**) also mentions a need for variability and argues that in some instances (e.g., operated rats), the capacity to vary behavior exists but is not used. In a later paper, with Schneirla (**149**), he again considers the tendency to vary behavior. Mowrer and Jones (**159**), explored variability as a function of the effort involved in the responses. In Cattell's formulation (**25**, p. 635), there is a "law of dispersion with excitement and deprivation. Continued stimulation of ergs, with deprivation of the goal, produces increasing variation in the stimuli to which attention is directed and increasing variation of response behavior (as well as introspectively, increased 'excitement')."

In his formal theory, Hull (**113**, pp. 304–321) gave explicit recognition to the problem of variability by his concept of behavioral oscillation. This oscillation was postulated to be specific to each reaction potential, i.e., the oscillations of different reaction potentials are asynchronous. He also called attention to other kinds of variability: "The 'constant' numerical values appearing in equations representing primary molar behavioral laws vary from species to species, from individual to individual, and from some physiological states to others in the same individual at different times, all quite apart from the factor of behavioral oscillation ($_SO_R$)" (**114**, p. 117). Subsequently, he considered oscillation in relation to conflict situations, behavioral inconsistency in evaluative choices, and alternation tendencies (**115**). Taylor (**207**, **208**) discussed the concept of behavioral oscillation and argued that it is related to strength of drive.

Statistical learning theory has recognized the problems of stimulus and response variability (Estes, **61**; Estes and Burke, **62**). Their formulations may lead to a model that can be fitted to the intraorganic determinants of response variability.

Psychophysics would seem at first thought to be an excellent source of data fitting our Type I model of variability, since many of its methods involve repetitions of the same stimulus. However, most of the research in this area is normative—it is concerned with general functions, not differences between individuals. Thus Guilford (**88**) neglects differences in intra-individual variability. Thurstone (**215**) points out that the slope of the psychometric function for the constant method indicates degree of sensitivity. Bevan and Dukes (**11**) show that smaller average errors are obtained in judging the distance of a more valued stimulus object, and Dukes and Bevan (**57**) found less variability of response in comparing two positively-valued objects than in comparing two negative objects or a positive and a negative object.

Accuracy of performance. Some of the first work (e.g., **210**) was on variability in accuracy of performance. For several studies of repeated testing, Thorndike (**211**) studied the distribution of the individual's deviation from his own average. Since this study utilized alternate forms of the tests, it does not fully fit the paradigm (cf. **213**).

Flügel (**71**) studied intrasession oscillation and day-to-day variability in continuous addition problems. Absolute oscillation was measured by the sum of changes between successive brief periods and absolute variability was based on deviations from the mean of the five sessions centered around a given day. Using both absolute and relative measures, he found that oscillations and variability were positively associated. He points out that in his data, oscillations usually represented dips from a relatively steady rate, rather than spurts of faster performance.

Hertzman (**107**) developed a variability score based on deviations from each subject's median on the Thurstone Substitution Test. For a group of highly variable subjects, the size of the intercorrelations between variability scores for different trials was a function of proximity in time. A similar but less marked trend was found in the low variability group.

Variability in hand-arm steadiness within one session was studied by Lovell (**142**). Three measures were used: average deviation, relative variability (average deviation divided by mean), and sum of successive differences between pairs of trials. For all three, substantial correlations were obtained between scores for two sessions a month apart. The values ranged from .49 to .74, with those for relative variability being the lowest. (Scores were computed for each of three time intervals: 1, 3, and 12 seconds.)

Jarrett (**122**) reports individual differences in proportions of responses changed on a multiple-choice test.

Output or magnitude of response. The well-known work of Dodge (**54**) employed a series of measures at different levels of neural integration. Darroch (**44**) reported daily fluctuations on a perseveration test (where the score is actually based on the difference between performances on two tasks). Johnson (**124**) found wide individual variation in finger pressure.

Marked individual differences in variability of judgments about a series of colors were reported by Hunt and Flannery (**117**). Variability increased with the number of colors to be judged and the number of categories used, but decreased as the number of repetitions mounted. David and Rabinowitz (**45**) developed an instability score based on changes in preference for the Szondi pictures. Our analysis of their data shows substantial correlations between instability scores for the eight "factors" in the test, the average intercorrelation being .42 for nurses and .62 for paranoid schizophrenics. (These are somewhat spurious: because it is a forced-choice test, a change in preference for one picture must produce a change in preference for some other picture.)

Several studies of the reliability of paper-and-pencil inventories have data on variability. Substantial internal consistencies were found by Lentz (**138**) for number of items changed and for change in total score, these two measures correlating .40. As we might expect, certain items contributed more changes than others, especially the more general and abstract ones. Glaser's criticism (**78**) that the relationship between level scores and change scores is a function of the range and design of the test may be applicable to these findings (cf. **163**, **173**).

Variability has also been noted by students of personality. The reliability of personality measurements is a function of the stability of the individual personality, according to Maller (**150**). Roshal (**185**) concludes that successful therapy is associated with gains in behavior variability which contributes to adaptability. Rapaport (**176**) observes that "variability of reactions" follows from "the multiple determination of psychic events." Saudek (**189**, p. 62) discusses changes of pressure in handwriting. Spontaneous changes in characteristics of responses within one Rorschach examination are discussed by Beck (**9**). In a study of changes in responses to a modified Rorschach test, Siipola *et al.* (**196**) found that almost half the responses were changed more or less. Allen *et al.* (**2**) report that two-thirds of the responses were changed. Gibby (**75**) studied the stability of intellectual variables for repeated Rorschach tests. Fluctuations from test to retest on the Bender-Gestalt are reported by Pascal and Suttell (**165**), who suggest that attitude may be the explanation. It seems reasonable to hypothesize that some variability can be accounted for by change in attitude or set; in fact, it is possible that differences between individuals with respect to variability may be a function of differences in the strength of a set and in the capacity or need to maintain it.

Other work. Several studies cannot be grouped into the categories used above. The variation in response pattern shown by mammals when confronted with an unsolvable task is more a characteristic of the individual than of the species, according to Hamilton (**96**).

Variability of the instrumental behavior in guinea pigs was studied by Muenzinger (**160**), who noted that plasticity remained even after one thousand trials in a puzzle box. Some animals would repeat one movement for a large number of trials and then shift to another one. In a subsequent study (Muenzinger, Koerner, and Irey, **161**), animals which were required to make a specific movement in order to escape showed variation in the required movement but also showed a greater number of accessory movements.

Guthrie and Horton (**90**) observed the behavior of cats in repeated escapes from a puzzle box. They concluded that the cat tends to repeat his previous response. Inspection of their photographic records, however, indicates marked differences between the cats in their degree of consistency.

Rimoldi mentions a relationship between mean reaction time and variation: "a small coefficient of variation and high speed tend to go together, while in slow individuals the variability may be either high or low" (**181**, p. 298). (Cf. also **136**, **146**.)

The studies discussed above have demonstrated variability in performance with respect to accuracy, magnitude, and rate, and have illustrated variability in instrumental acts. The evidence justifies the authors' view that significant components of such variability can be isolated and interpreted.

STUDIES OF SYSTEMATIC VARIABILITY

In this section we shall consider investigations using repetitions of the same stimulus in the same situation where the responses show no monotonic trend but where some regularity or order is noted (Type II variability). Thus studies involving improvement and deterioration of the quality, strength, or rate of response (e.g., due to learning or fatigue) will be omitted, but we shall include data that are viewed as a stationary time series, where the starting point is immaterial. We shall omit references to phenomena showing fixed cycles—such as diurnal fluctuations.

Most examples of systematic variability can be classified as reactive variability—the subject responds not only to the re-presented stimulus but also to the previous presentation or to his previous response to it. In a discussion of difficulties in using repeated measurements, Smith (**198**) points out that successive responses depend more or less on preceding stimulus-response situations. The problem is made more complex by the fact that the extent to which a response is influenced by preceding responses varies throughout the series, perhaps even from one response to the next. Johnson (according to London, **141**) also argues that the state of the organism is altered by the previous response. The non-independence of successive responses in measuring visual threshold was reported and analyzed by Verplanck *et al.* (**221**). Wertheimer (**226**) corroborated this result and also found significant interaction variance between days and subjects.

Serial Order

Variability measured by differences between successive responses falls somewhere between the major Type I and Type II classifications. Since the order of the responses is considered, it does not meet a rigorous interpretation of the first category. On the other hand, it does not imply the degree of regularity found in the reactive or periodic forms of systematic variability.

Philip (**171**) had twelve subjects tap alternately on two plates at maximum speed until exhausted. In seven subjects he found evidence for periodicity, the periods being about 50 minutes in length. The periodicity was measured by three methods—one of which was the nonrandomness of distribution of the most frequently occurring intervals between taps. He distinguished between periodicity (which may be an efficient mode of operation because it provides opportunity for recovery) and variability due to spurts and to the effects of distractions.

Abelson (**1**) applied spectral analysis to performance on repetitive tasks. This technique assesses the components of the output curve without assuming periodicity. Using an analogy from engineering, he asked whether individuals showed differences in degree and type of "out-of-controlness." (His study does not completely fit our paradigm since his stimuli had some variety, but he assumed with some justification that the analysis could disregard these differences.)

The task was a perceptual-motor one: making 100 jabs at each of five targets. The technique permitted recording the performance with the subject having only minimal knowledge of results.

Abelson obtained repeat measurements on some subjects. As one might expect, the repeat reliability of variability measures for the first task was much lower than for the succeeding four tasks. While the variance for each individual and the value derived from spectral analysis were uncorrelated with each other, each showed stability from test to retest. It may be worth noting that the retested group were volunteers; as a group, they showed on their first test a higher level and a greater scatter of their individual variances than did the group not volunteering for retest.

The Conceptualization of Systematic Variability

The most prominent figure in the history of work on variability is Raymond Dodge. He was one of the few people to study the general problem. He pointed out that variability could be viewed as accidental deviations from some abstracted "true" measure, but that such an approach with its emphasis upon invariants might overlook significant aspects of behavior (**52**):

> The psychophysical organism is in a perpetual state of flux. . . . Moreover, the neuromuscular consequences of two successive instances of stimulation with physically similar stimuli vary not only according to the momentary conditions of the organism and its psychophysical set, but also according to inner reactions and inhibitions. As is well known, the repetition of identical stimuli may not evoke the same reaction in successive instances (**55**, p. 5).

Dodge first published his two laws of relative fatigue in 1917. "Within physiological limits, all fatigue decrement in the results of work is relative to the intensity of the stimulus" (**51**, p. 102). "In any complex of competing tendencies, the relatively greater fatigue of one tendency will tend to eliminate it from the competition in favor of the less fatigued tendencies" (**51**, p. 105). Recognizing the phenomenon of avoidance of repetition of a response, he suggested the possibility of making use of it to study the nature of inner stimuli. He points out that relative fatigue should be viewed positively, as a conserving mechanism helping to prevent exhaustion.

Dodge (**53**) collected voluminous data on several levels of neuromuscular processes, for one subject. The records extended over two years and were taken at various times of day, etc. Unfortunately, Dodge attempted to interpret his data by a series of analogical constructs derived from contemporaneous neurophysiology. Thus he extended the neural concept of refractory phase to reduced responsiveness persisting for minutes and even years.

Solomon (**201**) notes that Hunter, observing a tendency to alternate (**118**), assumed it was innate (**119**). Hull (**113**) utilized "reactive inhibition" in his system, and Underwood (**220**) makes this concept the basis of his explanation of alternation. Rothkopf and Zeaman (**186**) suggest that the alternation tendency has a large "response" component and two small "place" components. Hebb (**103**, p. 228) has also suggested an interpretation of this phenomenon.

Two papers by Glanzer (**76, 77**) promulgate the provocative concept of stimulus satiation and give experimental evidence to support it. His basic assumption is that with continued exposure to the same stimuli in the same environment, the organism becomes less active. Every moment the organism perceives a stimulating object, there develops a quantity of stimulus satiation to it. Glanzer maintains that the same principle holds for all multiple-choice and free response situations.

To support his theory, Glanzer provides a telling critique of response theories which are based on the avoidance of the last response. Most of the studies in this area can be interpreted by either theory because it is difficult to determine whether the stimulus or the response is being avoided. There are few crucial experiments other than Glanzer's own work. Although Glanzer does not attempt to do so, it may be possible to adapt his theory to apply to verbal responses. While the nonrepetition of a recent verbal response might seem to be a case of reactive inhibition, it could be explained as the avoidance of stimulus satiation. If the organism is repeatedly exposed to the same stimulus in the same environment, and if it is forced to respond (i.e., if it cannot become less active), it will introduce variety in its responses. It is obvious that nonrepetition of response will reduce boredom.

In this connection, an early study by Robinson and Bills (**182**) is pertinent. On the basis of introspective reports, they conclude that the rapid repetition of homogeneous responses succeeds best without full attention; subjects were most efficient when they were having "concrete fantasies."

The Nonrepetition of the Preceding Response

A large number of papers deal with the relationship between successive pairs of responses. In a situation with repetitions of the same stimulus and with two or more comparable alternatives, is there a tendency for a response to be different from the preceding one?

Motor responses. Lewin (**139**) reports an experiment where, with some pressure on them to continue as long as possible, subjects drew repeated moonfaces. While morons either drew them, or broke off to pause or to do something else, normal subjects used secondary activities and other means to keep going.

Most of the work on alternation of motor responses has been done with rats (see reviews in **115** and **201**). Yoshioka (**230**) reported that only a small proportion of his animals distributed their responses evenly between two equally good paths to a goal, and these showed alternation tendencies. MacGillivray and Stone (**143**) concluded that the tendency to alternate responses is much stronger than the tendency to repeat.

Wingfield and Dennis (**227**) report 91 per cent alternation for rats given two trials a day but only 68 per cent for six trials a day. Dennis later (**48**) concluded that alternation effects were due not to a tendency to alternate direction but to a tendency to avoid a specific pathway already taken. However, this latter tendency was not found in mazes with more than two choice points.

Leeper (**137**) trained rats to choose one path when hungry and the alternative path when thirsty. In considering responses on the second trial on each day, he invoked the idea of a "systematic tendency to variability."

Further light was thrown upon spontaneous alternation by Heathers (**102**) who found that the percentage of alternation decreased as the time between trials increased from 15 to 120 seconds; when the interval was 15 minutes, alternation disappeared. Weitz and Wakeman (**225**) reported that alternation decreased to a minimum at intervals of 40 to 50 seconds and then rose with longer intervals. Alternation was noted by Riley and Shapiro (**180**) for trials 25 seconds apart (but not for trials 5 minutes apart). This tendency declined as the trials continued.

After Solomon's extensive survey of "The Influence of Work on Behavior," he concludes that "work acts to produce negative motivation" (**201**, p. 35). Solomon (**200**) was able to increase the percentage of alternation to a maximum of 90 per cent by making rats go up an inclined ramp (according to Zeaman and House, **232**). Rothkopf and Zeaman (**186**) report that alternation was increased by more forced trials and also that it increased as the series of daily trials continued. In a very recent study using meal worms (**85**), Grosslight and Ticknor found that alternation is increased by a forced choice and by the summated effect of two preceding turns, but is decreased by a longer distance from the last choice point.

A series of experiments on spontaneous alternation has been reported by Montgomery (**152**, **153**, **154**, **155**, **156**). He believes that spontaneous alternation may be a special case of exploratory tendency. The exploratory tendency is reduced by the exposure to a place (e.g., an arm of a maze), thus leading to an explanation based on place avoidance, not response avoidance. This point of view resembles Glanzer's concept of stimulus satiation (**77**).

Montgomery concluded that "in simple maze-situations, amount of exploratory behavior decreases as time of exploration increases, and increases proportionally as the area available for exploration increases" (**154**, p. 584). The same tendency reappeared on each day's trial (**155**). In both studies, the proportion of alternations in sequences of locomotor responses was above chance.

In another study (**153**), the percentage of alternation declined as the interval between trials was lengthened. Within each block of ten trials, the alternation rate was constant from the first half to the second. The effort required to press the bar did not affect alternation. The discrepancy between this finding and Solomon's (see above) may be due to any of several factors, such as the kind of work involved.

One group of papers has emphasized the absence of variability. Hamilton and Ellis (**98**) concluded that normal rats showed persistency in seeking a goal, but variability in their behavioral activities, whereas operated rats showed behavioral constancy—a relatively fixed sequence of behavioral acts functioning as a unit. Hamilton and Krechevsky (**99**) demonstrated that a shock administered just before the rat reached the choice point was associated with a reduction in variability and a fixation on one choice or the other. In several experiments, the reduction of variability by shock and by conflict was found by Everall (**63**) who noted a tendency for the same rats to perseverate under the different conditions. Krechevsky published three papers on "Brain-Mechanisms and Variability." In the first (**132**), he found that normal rats used more different paths to reach a goal and shifted the path used more often than did operated rats. In the latter group the size of the lesion was negatively related to the number of paths used. Since he did not count paths on which rats made errors, part of these findings may be a result of the fact that the operated rats made six times as many errors as the normal rats. In the second study (**133**), normal rats preferred a longer path of varying shape to an alternative, shorter path of fixed length, but operated rats did not. On the other hand, in an experiment where the varying path was much shorter, no significant difference was found between the two groups: both chose the varying, shorter one only slightly more than half the time (**134**). Krechevsky concluded that operated rats can show variability of response when there is a large difference in the efficiency of the alternatives. However, these are group averages: the operated rats showed a greater tendency to perseveration, to repeat their last response, especially in the second experiment (cf. **147**).

Verbal responses. Most studies of patterning in the response sequences of human subjects have used guesses or psychophysical judgments. Since the literature in this area has been reviewed recently (**194**, **201**), we shall not attempt a complete coverage here.

The nonrandom patterning of verbal responses was noted and discussed by Thorndike (**212**) and by Dodge (**54**). In his intensive study of one subject, Dodge (**53**) included vocal reaction time to a series of words presented in random order. (In written associations, Telford [**209**] found less repetition with shorter time intervals between responses.) In a study of speed of naming colors Bills (**12**) believed he had evidence for blocking. Abelson (**1**) reanalyzed Bills' data and found no rhythmicity.

In a series of five choices between two alternatives, Goodfellow (**83**) found a tendency to avoid symmetrical patterning. Skinner (**197**) argued that Goodfellow's data could best be explained as tendency to alternate which is strengthened if the two preceding responses show repetition, and is weakened when they show alternation. Solomon (**202**) found that alternation tendencies in guessing were unaffected by the interval between guesses or by the effort required to record their guesses.

In a psychophysical experiment using the method of constant stimuli, a tendency to avoid repeating the preceding judgment (especially the judgment of equal) was noted many years ago by Fernberger (**69**). A similar tendency was reported by Turner (**218**), by Arons and Irwin (**6**), and by Irwin and Preston (**121**). (Cf. **198**, discussed above.)

Day (**46**, as reported by Abelson, **1**) studied patterns in differential threshold responses to auditory stimuli and found long successions of correct discriminations and of failures to detect differences. From his reanalysis, Abelson suggests the interpretation that the probability of detecting a difference is greater when a preceding difference has been heard and less when the preceding change was not heard.

Other Work on Systematic Variability

There are several related areas which will be mentioned but will not be comprehensively reviewed.

Periodicity. The search for periodicity in performance has a long history. As early as 1905, Seashore and Kent (**193**) reported periodic waves in continuous mental work. Sarvis (**188**) concluded that rhythms were ascertainable in the time required for continuously tracing mazes blindfolded. However, he did not solve the problem of objective tests for rhythmicity: he held that "prolonged experience" was necessary in making judgments about the presence of rhythms.

Philpott has been concerned with this problem for more than twenty years. In an early monograph (**172**), he provided a history of older work on curves of output and attempted to demonstrate geometric periodicity. He has recently sought to relate his psychological constants to physical constants. Among those unconvinced by Philpott's arguments and concerned about his failure to utilize statistical tests is Richardson (**179**) who tested a work curve considered representative by Philpott and concluded that its spikiness might be random.

Oscillation. Spearman (**203**) provided an extended discussion of oscillations in efficiency. He held that these are manifested in fluctuations of minimal sensory impressions, in fluctuations of mental output, and in rivalry such as reversible perspectives. He concluded that there is a general oscillation factor which cannot be explained by *g* or by perseveration. Tussing (**219**) found that the rates of fluctuation of four illusions increase with physical fatigue.

Flügel's work on oscillation is discussed above. (Cf. also the section on Serial Order.)

Vacillation. Another more or less tangential topic is vacillation in conflict situations. A systematic discussion is provided by Miller (**151**). His paradigm is based on the relationship between distance and strength of response tendency. When an approach gradient and an avoidance gradient cross, vacillation of response occurs at the intersection. More vacillation is found in the stable equilibrium produced by two avoidance gradients than in the unstable equilibrium of two approach gradients. More complex situations are also considered.

Systematic covariation. The most extensive work on covariation is that of Cattell. He and his colleagues have reported several studies (**27, 28, 29**) of *P* technique, the correlation of measures on the same individual made on a number of successive occasions. Fluctuation of attitude seems to be associated with emotionality (low maturity). A central consideration in this work is the relative range of fluctuations in different functions and the influence of that range on the obtained relationships.

Holt (**110**), following a suggestion from Horn (**111**), correlated self-ratings and Szondi scores for one subject over twelve trials. A large number of self-rating items had high correlations over trials with Szondi factor *m*.

Some covariation tendency for psychometric tasks was noted by Asch (**7**); the scores of each individual tended to vary together over time, suggesting a general efficiency factor. Our analyses of his data reveal no relationship between the extents of total variability on three different tasks.

We have intentionally omitted any consideration of mood swings in both normal and pathological subjects.

In summary, there is considerable evidence of the existence of more or less systematic variability. The most common finding is that among equally efficient response alternatives, a given response will differ from the immediately preceding one, and even from other very recent ones. This is particularly true where the response is a choice of means to the same goal. It also occurs among choices where knowledge of the accuracy of the choice is not available to the subject. There is no general agreement on the explanation for this reactive variability or other systematic variability.

Type II variability can be viewed as a special case of Type I variability. Systematic variability occurs only when the alternative responses have comparable probabilities of occurrence and when the successive presentations of the stimulus are relatively close together in time.

CHANGE OF RESPONSE WITH CHANGE IN STIMULUS OR SITUATION

In this section, we shall consider Type III variability, which we defined earlier as variability in response with variation in the stimulus or in the situation. While this class is obviously very large, some of the research in this area has implications for our central problem. Of course, all variability is of this type if one takes Guthrie's position (**89**) that the exact situation is never repeated.

Change in Situation Primary

Variability in this category is basically the difference between two responses to the same stimulus in two different situations.

A great deal of psychological research utilizes the basic design of comparing responses under two different experimental conditions. The interest is generally in the mean change in total score or average performance rather than in absolute change on separate items. Furthermore, the emphasis is usually on the group change, rather than on the distribution of individual change scores. For example, Johnston (**125**) reports low but generally positive relationships between measures of adjustment and relative gain from test to retest when the retest was administered under stressful conditions. An illustration of a change score based on absolute changes in responses to items is Brownfain's study (**20**) of the stability of the self concept under varied instructions.

Change in Stimulus Primary

While the concept of variability is closely related to the concept of rigidity, they can be distinguished by two differences in emphasis. In the usual paradigm for studying rigidity, the subject is presented with first one stimulus and then another, objectively different one, with the general situation more or less constant. The focus is on the extent to which the subject fails to change his response, i.e., upon the degree of invariability. In approaching variability, on the other hand, the more typical study keeps both the stimulus and the situation constant and measures the degree of variability, which may be excessively small *or* large. The second difference is in normative emphasis. Rigidity is conceived to be maladaptive, and hence as the opposite of adaptive, appropriate response tendencies. In studying variability, the adaptiveness of the change in response is secondary. Change may or may not be considered desirable. On the one hand, fluctuations in quality or accuracy may be viewed as undesirable; on the other hand, the tendency to make use of more than one of several equally efficient alternative responses may be beneficial because it sustains efficiency by preventing boredom. Thus the problem of variability encompasses and is more general than the problem of rigidity.

Krechevsky and Honzik (**135**) found that individual rats which consistently chose the shorter of two paths in a maze could change, when the paths were interchanged, as rapidly as those showing more variability on the first problem (cf. also **131**). In a study of human problem solving, Guetzkow (**86**) concluded that there are two distinct factors in set: susceptibility to set (tendency to acquire a set readily) and ability to surmount an acquired set. Churchman (**33**, p. 240) conceives of personality "as the measure (or measures) of typical inefficiency an individual displays in problem-solving," due to his failure to drop an old method or his tendency to change to a new, less efficient method.

Cattell and Tiner (**30**) have provided a review of the literature on perseveration and a factorial study of structural rigidity. Most of their tests involved capacity to perceive stimuli in new ways, not tendency to respond differently. A further analysis of structural rigidity has been made by Cattell and Winder (**31**), who distinguish between fluctuations in goals and fluctuations in goal paths. Kleemeier and Dudek (**129**) sought a general flexibility factor but failed to find it. However, their flexibility tests had little homogeneity.

A major monograph on rigidity is

that by Fisher (**70**). For him, a measure of rigidity is the number of equivalent alternatives which the subject demonstrate, in some behavioral way, that he can utilize. Once again, the emphasis is upon modes of response which are possible for the subject, not upon variability in actual responses. Thus he used the number of objects liked in a given set of stimuli, the number of possible alternatives accepted by the subject, etc. Several of his suggestions concerning rigidity may have analogues in variability: for example, his discussion of individual consistency of rigidity manifested on tasks of a given difficulty level as related to adjustment, and his distinction between inner rigidity when self-esteem is threatened and peripheral rigidity when not emotionally threatened.

RELATIONSHIPS BETWEEN VARIABILITY AND OTHER VARIABLES

Some Experimental Variables Related to Variability

In this section, we shall review evidence concerning the relationships between variability and each of several kinds of variables. Some studies involving intra-organic conditions will be discussed in this first part.

Brain lesions appear to be associated with reduced variability. Halstead (**95**) noted that the average deviation of trials on critical fusion frequency trials is lower for frontal lobectomy cases. Reference has been made earlier to several pertinent studies (**98, 130, 132, 133, 134**).

The effect of shock on variability has been studied by several people (**63, 99, 162, 187**). Fairlie (**66**) discovered that shock at the choice point produced more fixation in rats entering the correct path than in rats entering an incorrect one. De Valois (**50**) offers the interpretation that fixation occurs when the rat cannot avoid the punishing shock. After discussing the studies (**126, 128, 164**) that show the fixation persisting after the shock is discontinued, he emphasizes Farber's finding (**67**) that the persisting fixation is caused by anxiety: when the anxiety state is removed, the fixation is ended.

De Valois' own experiment (**50**) is a major contribution to this area. He showed that in rats, more intense motivation is associated with lower variability, for both an approach motive (thirst) and an avoidance motive (shock). Increasing the motivation decreased the variability, and decreasing the motivation increased the variability. The latter effect was demonstrated when the original motivation was moderately high but not when it was very strong. De Valois does not accept Maier's position, which includes the principle (**148**, p. 159) that a problem situation produces stereotyped behavior in a frustrated individual but variable behavior in a motivated one. De Valois' results agree with Elliott's finding (**60**) that increased hunger lowered variability but that the rats remained at a low level of variability subsequently when they were less hungry.

In human subjects, increase in motivation may increase variability. Deese and Lazarus (**47**) obtained greater variability of performance on a Rotary Pursuit Test by making the task more important and by inducing failure stress. These factors also increased the interindividual differences in variability. A measure of the variability of reaction time for binocular fusion under conditions involving emotional stress may be useful in selecting emotionally stable people (Brown University, **19**). Using only 14 subjects, Baker and

Harris (**8**) found Rorschach correlates with an increase in variability of speech intensity under stress (threat of shock). It should be noted that these stresses used with human subjects are disruptive in part because no positive adaptation is possible.

The picture at lower levels of motivation is less clear. J. G. Taylor (**207**) suggested that strength of drive is related to amount of spontaneous activity, one aspect of which is behavioral oscillation. Solomon (**201**) reported an increase in alternation with increase in effort required. On the other hand, with sets of two equal paths, Mowrer (**158**, Ch. 6) working with Orbison found that the longer the route, the less the vacillation. This finding may be a function of increased time between trials and/or reduced amounts of stimulus satiation effects (cf. Glanzer, **77**); the increase in energy expenditure may be unimportant here.

In an experiment by Goodman, Moyer, and Bunch (**84**), rats were exposed to electric shock, air blast, or food deprivation before being dropped into water. These various conditions did not produce differences in variability of alley chosen to get out of the water, perhaps because the conditions were not sufficiently stressful.

Several studies have noted changes in variability, both within one occasion and over occasions (**1**, **94**, **159**, **207**).

Personality as Related to Variability

Personality traits. Hall (**94**) found that rats showing more variability among five alleys of equal length also showed greater emotionality (as measured by defecation in the apparatus) and took somewhat more time per run. Emotionality might be construed as an index of unreduced tension.

Frenkel-Brunswik (**72**) reported some provocative relationships between trait ratings and fluctuations in ratings made semiannually.

Cattell (**23**) confirmed his prediction that degree of change in sentiments and attitudes would be negatively related to w, the general character factor, as rated by peers. He explained the lower correlations of his measure of fluctuation with extroversion, emotionality, and mood swings in terms of the relationships between these variables and w. Fairly consistent results were obtained for children and for adults, using intervals of either one day or one month between the administrations of the inventories. The w factor has its highest correlations with small change in "deeper sentiments." Similarly, Cummings (**43**) found that variability on self-ratings correlated negatively with persistence, w, and introversion. Subjects high on this kind of variability were rated by others as original, imaginative, and talkative; low subjects were regarded as conventional, thorough, and pugnacious. Walton (**222**) demonstrated a relationship between steadiness of character and low oscillation on motor and cognitive tests.

From a study of weekly retests on the MMPI, Layton (**136**) concluded that variation is a function of the individual, not of his score relative to the group. Variation on single scales had no consistent relationship to mean score although some trends were in a plausible direction.

Rosenzweig and Mirmow (**184**) found that degree of socialization was associated with trends on the P-F test, a trend being a shift from extrapunitiveness to impunitiveness during the testing. (This type of change is presumably due to the subject's reaction to his previous responses.)

Two papers (**163**, **173**) correlated changes in questionnaire responses with test scores on adjustment. Glaser, however, has done a more thorough study (**78**) demonstrating that the correlations between level scores and consistency are zero for tests with ranges appropriate for the group tested. This criticism does not apply to Weber's study (**224**) in which variability of speed on psychometric tests was related to emotionality and submissiveness (as measured by questionnaires.)

Several factor analyses including variability measures have been made. Brogden (**17**) found nonvariability among 30 trials on addition had an appreciable loading on a factor involving ability to work steadily and to resist distraction. In his synthesis of factor studies using objective personality tests, Cattell (**26**) notes several factors on which variability measures have loadings. Connor (**34**), using measures of both daily variation and immediate variation, found little relationship between temperament and variability measures from which ability had been partialed out.

Personality integration. In several sources, we find the suggestion that personality integration is negatively related to variability (cf. **23** and **125**, discussed above; also **109**, **204**).

A provocative paper by Smith and Klein (**199**) indicates a relationship between variability of speed on the Stroop color-name test and a type of adaptability which is accompanied by lability of control.

In their review of the Szondi test, Borstelmann and Klopfer (**16**) evaluate the proposition that the picture categories in which changes in preference patterns occur represent the more unstable areas of personality, whereas categories representative of stable, basic need systems will show little change over time in selection pattern (cf. **49**). Noting that David and Rabinowitz (**45**) found greater changes in choices for schizophrenics than for student nurses, Borstelmann and Klopfer conclude that variability in Szondi Test behavior occurs in the records of normal subjects and, to a greater extent, in the response of pathological subjects. For both groups, the "variability seems to be pervasive and not differential among test categories" (**16**, p. 124).

Personality types and diagnostic categories. Pauli utilized a variety of measures in his analysis of the curve of performance for continuous additions and subtractions (**166**). In a second paper (**167**), he concludes that to assess character and aptitude, the total number of items tried and the total number correct are sufficient; the range and average deviation are unecessary. Susukita (**205**), using Pauli's methods, found he could distinguish between two Japanese character types: the inner-integrative or rigid had lower variation in performance than the outer-integrative or labile. Variation was not related to age. Andō (**4**), however, concluded that variability on psychophysical tasks could not be used for character diagnosis. Using instructions to sustain pressure on a dynamometer at a constant level, Eden (**59**) found that the graph of pressure varied for the different Jaensch types. In general, the **J** types had smooth curves while those for the **S** types were irregular.

The variability of performance of normal and neuropsychiatric groups on various tasks has been compared by Eysenck and his colleagues (**64**) and by Roseman (**183**).

The question whether psychotics, especially schizophrenics, are more variable than normals has been examined in several papers. The earliest study was made by Gatewood (**74**) who found that dementia praecox patients were more irregular than normals. He suggested that the difference was due to the patients' poor attention and their lack of "thought control." Seeking to determine whether schizophrenics were more variable because of defects in "psychological government," Hunt (**116**) obtained self-reports from his subjects. He found schizophrenics were more variable and that, within that group, output was significantly related to type of preparatory set taken by the subject. The greater variability in output was associated with greater variability in set toward the task. Two other possible governing factors had little or no independent effect.

A similar study was carried out by Huston, Shakow, and Riggs (**120**). Their schizophrenics had higher means and intra-individual variability in reaction time than their control group. Since the schizophrenics showed lower means and variability in their second and third testing sessions than in their first, the authors suggest that schizophrenics may have a slower rate of adaption. Cooperation and variability were negatively related, but even the more cooperative patients were more variable than the controls. A second experiment tended to confirm the hypothesis that schizophrenic patients do not attain as high a level of preparation or set, show more variability in height of preparation, and do not maintain their best level of preparation as consistently as do normal subjects. An earlier study by Shakow and Huston (**195**) had shown greater variability for both schizophrenics and manic-depressives than for normals on speed of alternate tapping.

Using tests similar to those of Allport and Vernon (**3**), Wulfeck (**229**) measured intra-individual consistency of performance by test-retest correlations between the second and third sessions. The average r for several tests was .81 for manic-depressives, .80 for normals, .75 for psychoneurotics, and .71 for schizophrenics. Schofield (**191, 192**) has examined the changes in MMPI responses following different therapies for normals, neurotics, and psychotics. Differences in the intra-individual variability of schizophrenics and controls with respect to physiological functioning are discussed by Hoskins and Jellinek (**112**).

There have been a number of papers on consistency of intelligence test performance (e.g., **104, 174, 175, 223**).

Variability or Variabilities?

With almost no exceptions, each of the studies reviewed in this paper has examined variability on only a single measure. We have no definitive evidence yet on the generality of variability. What is the factor structure of variability scores? Is there one general factor? Are there many common factors?

There is reason to believe that variability will turn out to be a function of the test stimuli and of the general situation. If so, the problem of the correlations between personality and variability becomes more complex. We may have to determine which personality traits are associated with each variability factor or measure.

SUMMARY AND CONCLUSIONS

Intra-individual response variability refers to the change in an individual's response from one time to the next. Variability under three broad classes of conditions has been considered:

Type I. Spontaneous or aperiodic

variability. The individual is presented with the same stimulus in the same situation at two points in time. It is assumed that the initial presentation and the initial response do not affect the second response, i.e., that the order of the responses is immaterial.

Type II. Systematic variability. Although the stimulus and the situation are again unchanged, the second response is influenced by the first presentation, by the first response, or by both. A primary example is reactive variability such as response alternation. Explicitly excluded from this type are (*a*) change showing a monotonic relationship with time (due to learning, fatigue, etc.) and (*b*) such periodic and cyclical phenomena as diurnal variation.

Type III. Change in response associated with change in stimulus or in situation. This comprehensive class includes all conditions not mentioned above. Only a few problems in this class, such as rigidity, have pertinence here.

Since the pure case, Type I, excludes all external determinants of change, variability must come from within the organism. We have, therefore, pointed out that fluctuations of intraorganic processes or states have been demonstrated or at least accepted by many writers.

Variability has been measured in many different ways. Furthermore, different variability scores may be based on different time intervals between the responses being compared. Measures of variability from one occasion have reasonable intercorrelations with each other and can have high stability over time, especially over short intervals. In general, we may expect the correlation between two of these measures to be a function of both the time interval between them and the relative homogeneity of the two tasks from which the scores were obtained. Such a score is itself much more subject to change over time than are conventional level scores. Variability measures appear to be related to set toward a task and to degree of adaptation to the stimulus situation.

Relatively little work has been done on Type I (spontaneous) variability, perhaps because it is rarely seen in its pure form and because it is difficult to elicit in experimental situations. Nevertheless, it has been recognized in almost all kinds of behavior.

Reactive variability (Type II) has been studied more thoroughly. Both motor responses in rats and verbal responses in humans show a tendency toward the nonrepetition of the preceding or very recent responses. It appears likely that the organism does not seek to avoid making the previous response but rather seeks to respond in such a way as to vary the total pattern of stimulation reaching it, including the stimulus produced directly or indirectly by its own response.

In examining some aspects of Type III variability, it was suggested that the concept of rigidity refers to variability which is restricted or reduced to a level considered to be maladaptive.

Variability is probably decreased by shock and by very strong motivation. Motivation at lower levels is less clearly or more complexly related to variation in response, e.g., experimental stress may increase variability.

The personality correlates of variability, if any, remain to be established definitively. It is likely that variability is negatively related to persistence and "character." Similarly, we cannot state with assurance the nature of the relationships with neurosis and psychosis.

This paper has reviewed many studies with more or less relevance to variability. While sections of the general area have been investigated, few sections have been systematically attacked. The phenomena require further intensive studies.

BIBLIOGRAPHY

1. Abelson, R. P. Spectral analysis and the study of individual differences in the performance of routine, repetitive tasks. Princeton: Educational Testing Service, 1953.
2. Allen, R. M., Manne, S. H., & Stiff, Margaret. The influence of color on the consistency of responses in the Rorschach test. *J. clin. Psychol.*, 1952, **8**, 97–98.
3. Allport, G. W., & Vernon, P. E. *Studies in expressive movement.* New York: Macmillan, 1933.
4. Andō, M. Abweichung und Variation (Deviation and variance.) *Tohaku Psychol. Folia*, 1942, **9**, 223–232.
5. Antonitis, J. J. Response variability in the white rat during conditioning, extinction, and reconditioning. *J. exp. Psychol.*, 1951, **42**, 273–281.
6. Arons, L., & Irwin, F. W. Equal weights and psychophysical judgments. *J. exp. Psychol.*, 1932, **15**, 733–756.
7. Asch, S. E. An experimental study of variability in learning. *Arch. Psychol.*, N. Y., 1932, **22**, No. 143.
8. Baker, L. M., & Harris, J. S. The validation of the Rorschach test results against laboratory behavior. *J. clin. Psychol.*, 1949, **5**, 161–164.
9. Beck, S. J. *Rorschach's Test.* Vol. 2. *A variety of personality pictures.* New York: Grune & Stratton, 1945.
10. Bertalanffy, L. v. *Problems of life.* New York: Wiley, 1952.
11. Bevan, W., Jr., & Dukes, W. F. Value and the Weber Constant in the perception of distance. *Amer. J. Psychol.*, 1951, **64**, 580–584.
12. Bills, A. G. Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230–245.
13. Bills, A. G. *General experimental psychology.* New York: Longmans, Green, 1934.
14. Bills, A. G. *The psychology of efficiency.* New York: Harper, 1943.
15. Blair, E. A., & Erlanger, J. A comparison of the characteristics of axons through their individual electric responses. *Amer. J. Physiol.*, 1933, **106**, 524–564.
16. Borstelmann, L. J., & Klopfer, W. G. The Szondi Test: a review and critical evaluation. *Psychol. Bull.*, 1953, **50**, 112–132.
17. Brogden, H. E. A factor analysis of forty character tests. *Psychol. Monogr.*, 1940, **52**, No. 3 (Whole No. 234).
18. Brown, W., & Thomson, G. H. *The essentials of mental measurement.* Cambridge: Cambridge Univer. Press, 1940.
19. Brown University. A method of determining reaction time of binocular fusion under conditions of stress. (1942; Publ. Bd. No. 155789) Washington, D. C.: U. S. Dept. Commerce, 1947.
20. Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.
21. Brunswik, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–273.
22. Butorin, V. I. Disorders of attention in psychoneurotics. *J. gen. Psychol.*, 1938, **18**, 235–251.
23. Cattell, R. B. Fluctuations of sentiments and attitudes as a measure of character integration. *Amer. J. Psychol.*, 1943, **56**, 195–216.
24. Cattell, R. B. *Description and measurement of personality.* Yonkers-on-Hudson: World Book Co., 1946.
25. Cattell, R. B. *Personality: a systematic theoretical and factual study.* New York: McGraw-Hill, 1950.
26. Cattell, R. B. The principal replicated factors in objective personality tests. *J. abnorm. soc. Psychol.*, 1955, **50**, 291–314.
27. Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, 1947, **12**, 267–288.
28. Cattell, R. B., & Cross, K. P. Comparison of the ergic and self-sentiment

structures found in dynamic traits by R- and P-techniques. *J. Pers.*, 1952, **21**, 250–271.

29. CATTELL, R. B., & LUBORSKY, L. *P*-technique demonstrated as a new clinical method for determining personality and symptom structure. *J. gen. Psychol.*, 1950, **42**, 3–24.
30. CATTELL, R. B., & TINER, L. G. The varieties of structural rigidity. *J. Pers.*, 1949, **17**, 321–341.
31. CATTELL, R. B., & WINDER, A. E. Structural rigidity in relation to learning theory and clinical psychology. *Psychol. Rev.*, 1952, **59**, 23–39.
32. CHOCHOLLE, R. Quelques remarques sur les variations et la variabilité des temps de reaction auditifs. *J. Psychol. norm. path.*, 1948, **41**, 345–358.
33. CHURCHMAN, C. W. *Theory of experimental inference.* New York: Macmillan, 1948.
34. CONNOR, D. V. Variability in mental test performance. Paper read at Brisbane Branch, Brit. Psychol. Soc., Brisbane, Australia, March, 1953.
35. COOMBS, C. H. A rationale for the measurement of traits in individuals. *Psychometrika*, 1948, **13**, 59–68.
36. COOMBS, C. H. The measurement of psychological traits, a research program. In Wilma T. Donahue, *et al.* (Eds.), *The measurement of student adjustment and achievement.* Ann Arbor: Univer. of Michigan Press, 1949. Pp. 227–238.
37. COOMBS, C. H. The concepts of reliability and homogeneity. *Educ. psychol. Measmt*, 1950, **10**, 43–56.
38. COOMBS, C. H. A theory of psychological scaling. *Univer. Mich. Engng. Res. Inst. Bull.*, 1951. No. 34.
39. COX, J. W. *Manual skill.* Cambridge: Cambridge Univer. Press, 1936.
40. CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1949.
41. CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
42. CROZIER, W. J., & HOAGLAND, H. The study of living organisms. In C. Murchison (Ed.), *Handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934. Pp. 39–49.
43. CUMMINGS, JEAN D. Variability of judgment and steadiness of character. *Brit. J. Psychol.*, 1939, **29**, 345–370.
44. DARROCH, J. An investigation into the degree of variation in the score of a motor perservation test. *Brit. J. Psychol.*, 1938, **28**, 248–262.
45. DAVID, H. P., & RABINOWITZ, W. The development of a Szondi Instability Score. *J. consult. Psychol.*, 1951, **15**, 334–336.
46. DAY, W. Stimulus interval as a determiner of serial patterns in threshold responses. Unpublished master's thesis, Univer. of Virginia, 1951.
47. DEESE, J., & LAZARUS, R. S. The effects of psychological stress upon perceptual-motor performance. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52-19.
48. DENNIS, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, **28**, 305–312.
49. DERI, SUSAN. *Introduction to the Szondi Test.* New York: Grune & Stratton, 1949.
50. DE VALOIS, R. L. The relation of different levels and kinds of motivation to variability of behavior. Unpublished doctor's dissertation, Univer. of Michigan, 1952.
51. DODGE, R. The laws of relative fatigue. *Psychol. Rev.*, 1917, **24**, 89–113.
52. DODGE, R. Problems of human variability. *Science*, 1924, **59**, 263–270.
53. DODGE, R. *Elementary conditions of human variability.* New York: Columbia Univer. Press, 1927.
54. DODGE, R. Note on Professor Thorndike's experiment. *Psychol. Rev.*, 1927, **34**, 237–240.
55. DODGE, R. *Conditions and consequences of human variability.* New Haven: Yale Univer. Press, 1931.
56. DUDEK, F. J. Concerning "reliability" of tests. *Educ. psychol. Measmt*, 1952, **12**, 293–299.
57. DUKES, W. F., & BEVAN, W., JR. Accentuation and response variability in the perception of personally relevant objects. *J. Pers.*, 1952, **20**, 457–465.
58. DUNLAP, J. W. Relationships between constancy of expressed preferences and certain other factors. *J. educ. Psychol.*, 1936, **27**, 521–526.
59. EDEN, H. Analyse einer einfachen ausseren Willenshandlung unter dem Gesichtpunkt der Integrationstypologie. *Z. Psychol.*, 1937, **141**, 198–240.
60. ELLIOTT, M. H. The effect of hunger on variability of performance. *Amer. J. Psychol.*, 1934, **46**, 107–112.
61. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

62. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
63. Everall, E. E. Perseveration in the rat. *J. comp. Psychol.*, 1935, **19**, 343–369.
64. Eysenck, H. J. *Dimensions of personality.* London: Routledge & Kegan Paul, 1947.
65. Fagin, H. An investigation of the nature and magnitude of the components of retest unreliability. Unpublished doctor's dissertation, New York Univer., 1950.
66. Fairlie, C. W. The effect of shock at the 'moment of choice' on the formation of a visual discrimination habit. *J. exp. Psychol.*, 1937, **21**, 662–669.
67. Farber, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
68. Ferguson, G. A. *The reliability of mental tests.* London: Univer. of London Press, 1941.
69. Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, **3**, 126–150.
70. Fisher, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
71. Flügel, J. C. Practice, fatigue, and oscillation. *Brit. J. Psychol. Monogr. Suppl.*, 1928, **4**, No. 13.
72. Frenkel-Brunswik, Else. Motivation and behavior. *Genet. Psychol. Monogr.*, 1942, **26**, 121–265.
73. Fryer, D. Variability in automatic mental performance with uniform intent. *J. appl. Psychol.*, 1937, **21**, 528–545.
74. Gatewood, L. C. An experimental study of dementia praecox. *Psychol. Monogr.*, 1909, **11**, No. 2 (Whole No. **45**).
75. Gibby, R. G. The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables. *J. proj. Tech.*, 1951, **15**, 3–25.
76. Glanzer, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.
77. Glanzer, M. Stimulus satiation: a construct to explain spontaneous alternation, variability, and exploratory behavior. *Psychol. Rev.*, 1953, **60**, 257–268.
78. Glaser, R. A methodological analysis of the inconsistency of response to test items. *Educ. psychol. Measmt*, 1949, **9**, 727–739.
79. Glaser, R. Multiple operation measurement. *Psychol. Rev.*, 1950, **57**, 241–252.
80. Glaser, R. The application of the concepts of multiple operation measurement to the response patterns on psychological tests. *Educ. psychol. Measmt*, 1951, **11**, 372–382.
81. Glaser, R. The reliability of inconsistency. *Educ. psychol. Measmt*, 1952, **12**, 60–64.
82. Goldstein, H. The biochemical variability of the individual in relation to personality and intelligence. *J. exp. Psychol.*, 1935, **18**, 348–371.
83. Goodfellow, L. D. A psychological interpretation of the results of the Zenith radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601–632.
84. Goodman, R. W., Moyer, K. E., & Bunch, M. E., Variability and behavior constancy in white rats. *J. comp. physiol. Psychol.*, 1952, **45**, 460–467.
85. Grosslight, J., & Ticknor, W. Variability and reactive inhibition in the meal worm as a function of determined turning sequences. *J. comp. physiol. Psychol.*, 1953, **46**, 35–38.
86. Guetzkow, H. An analysis of the operation of set in problem-solving behavior. *J. gen. Psychol.*, 1951, **45**, 219–244.
87. Guilford, J. P. Fluctuations of attention with weak visual stimuli. *Amer. J. Psychol.*, 1927, **38**, 534–583.
88. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
89. Guthrie, E. R. Personality in terms of associative learning. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 49–68.
90. Guthrie, E. R., & Horton, G. P. *Cats in a puzzle box.* New York: Rinehart, 1946.
91. Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255–282.
92. Guttman, L. The test-retest reliability of qualitative data. *Psychometrika*, 1946, **11**, 81–95.
93. Guttman, L. Problems of reliability. In S. A. Stouffer, *et al.* (Eds.), *Measurement and prediction.* (Vol. IV of *Studies in social psychology in World War II.*) Princeton, N. J.: Princeton Univer. Press, 1950. Pp. 302–311.
94. Hall, C. S. Emotional behavior in the rat: IV. The relationship between emo-

tionality and stereotyping of behavior. *J. comp. Psychol.*, 1937, **24**, 369–375.

95. HALSTEAD, W. C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
96. HAMILTON, G. V. A study of perseverance reactions in primates and rodents. *Behav. Monogr.*, 1916, **3**, No. 2.
97. HAMILTON, H. C. The effect of incentives on accuracy of discrimination measured on the Galton bar. *Arch. Psychol.*, 1929, **16**, No. 103.
98. HAMILTON, J. A., & ELLIS, W. D. Persistence and behavior constancy. *J. genet. Psychol.*, 1933, **42**, 140–153.
99. HAMILTON, J. A., & KRECHEVSKY, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, **16**, 237–253.
100. HAMMETT, F. S. Observations on the relation between emotional and metabolic stability. *Amer. J. Physiol.*, 1920, **53**, 307–311.
101. HARMON, F. L. The reliability of metabolism measurements by the closed circuit method. *J. appl. Physiol.*, 1953, **12**, 773–778.
102. HEATHERS, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10**, 359–380.
103. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
104. HEIM, A., & WALLACE, J. The effects of repeatedly retesting the group on the same intelligence test. II. High grade mental defectives. *Quart. J. exp. Psychol.*, 1950, **2**, 19–32.
105. HERON, W., BEXTON, W. H., & HEBB, D. O. Cognitive effects of a decreased variation in the sensory environment. *Amer. Psychologist,* 1953, **8**, 366. (Abstract)
106. HERRINGTON, L. P. The relation of physiological and social indices of activity level. In Q. McNemar & Maud A. Merrill (Eds.), *Studies in personality.* New York: McGraw-Hill, 1942. Pp. 125–146.
107. HERTZMAN, M. The influence of the individual's variability on the organization of performance. *J. gen. Psychol.*, 1939, **20**, 3–24.
108. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
109. HOLLINGWORTH, H. L. Correlations of achievement within an individual. *J. exp. Psychol.*, 1925, **8**, 190–208.
110. HOLT, R. R. An approach to the validation of the Szondi Test through a systematic study of unreliability. *J. proj. Tech.*, 1950, **14**, 435–444.
111. HORN, D. Intra-individual variability in the study of personality. *J. clin. Psychol.*, 1950, **6**, 43–47.
112. HOSKINS, R. G., & JELLINEK, E. M. The schizophrenic personality with special regard to psychologic and organic concomitants. *Proc. Ass. Res. nerv. ment. Dis.*, 1933, **14**, 211–233.
113. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
114. HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
115. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
116. HUNT, J. McV. Psychological government and the high variability of schizophrenic patients. *Amer. J. Psychol.*, 1936, **48**, 64–81.
117. HUNT, W. A., & FLANNERY, J. Variability in the affective judgment. *Amer. J. Psychol.*, 1938, **51**, 507–513.
118. HUNTER, W. S. The auditory sensitivity of the white rat. *J. anim. Behav.*, 1914, **4**, 215–222.
119. HUNTER, W. S. The temporal maze and kinaesthetic sensory processes in the white rat. *Psychobiology*, 1920, **2**, 1–17.
120. HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16**, 39–82.
121. IRWIN, F. W., & PRESTON, M. G. Avoidance of repetition of judgments across sense modalities. *J. exp. Psychol.*, 1937, **21**, 511–520.
122. JARRETT, R. F. The extra-chance nature of changes in student's responses to objective test-items. *J. gen. Psychol.*, 1948, **38**, 243–250.
123. JELLINEK, E. M. The function of biometric methodology in pyschiatric research. In *Mental health*, Publication of the AAAS. Lancaster: Science Press, 1939. Pp. 48–59.
124. JOHNSON, B. Changes in muscular tension in coordinated hand movements. *J. exp. Psychol.*, 1928, **11**, 329–341.
125. JOHNSTON, J. J. Relationship between emotional adjustment and change scores on speed tests. Unpublished doctor's dissertation, Univer. of Southern California, 1948.
126. KLEE, J. B. The relation of frustration and motivation to the production of abnormal fixations in the rat. *Psychol. Monogr.*, 1944, **56**, No. 4 (Whole No. 257).
127. KLEEMEIER, L. B., & KLEEMEIER, R. W. Effects of benzedrine sulphate (am-

phetamine) on psychomotor performance. *Amer. J. Psychol.*, 1947, **60**, 89–100.

128. Kleemeier, R. W. Fixation and regession in the rat. *Psychol. Monogr.*, 1942, **54**, No. 4 (Whole No. 246).
129. Kleemeier, R. W., & Dudek, F. J. A factorial investigation of flexibility. *Educ. psychol. Measmt*, 1950, **10**, 107–118.
130. Klein, G. S., & Krech, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118–148.
131. Krech, D. Dynamic systems as open neurological systems. *Psychol. Rev.*, 1950, **57**, 345–361.
132. Krechevsky, I. Brain-mechanisms and variability. I. Variability within a means-end-readiness. *J. comp. Psychol.*, 1937, **23**, 121–138.
133. Krechevsky, I. Brain-mechanisms and variability. II. Variability where no learning is involved. *J. comp. Psychol.*, 1937, **23**, 139–159.
134. Krechevsky, I. Brain-mechanisms and variability. III. Limitations of the effect of cortical injury upon variability. *J. comp. Psychol.*, 1937, **23**, 351–364.
135. Krechevsky, I., & Honzik, C. H. Fixation in the rat. *Univer. Calif. Pub. Psychol.*, 1932, **6**, 13–26.
136. Layton, W. L. The variability of individuals' scores upon successive testings on the Minnesota Multiphasic Personality Inventory. *Amer. Psychologist*, 1952, **7**, 384. (Abstract)
137. Leeper, R. The role of motivation in learning: a study of the phenomena of differential motivational control in the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
138. Lentz, T. F., Jr. The reliability of opinionnaire technique studied intensively by the retest method. *J. soc. Psychol.*, 1934, **5**, 338–364.
139. Lewin, K. *A dynamic theory of personality.* (Trans. by D. K. Adams and K. E. Zener.) New York: McGraw-Hill, 1935.
140. Loeb, L. *The biological basis of individuality.* Springfield, Ill.: C. C. Thomas, 1945.
141. London, I. D. Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena. *Psychol. Rev.*, 1946, **53**, 170–188.
142. Lovell, C. A study of personal variation in hand-arm steadiness. *Amer. J. Psychol.*, 1941, **54**, 230–236.
143. MacGillivray, M. E., & Stone, C. P. Suggestions toward an explanation of systematic errors made by albino rats in the multiple light discrimination apparatus. *J. genet. Psychol.*, 1930, **38**, 484–489.
144. McNemar, Olga W. The ordering of individuals in critical flicker frequency under different measurement conditions. *J. Psychol.*, 1951, **32**, 3–24.
145. McReynolds, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, **46**, 131–141.
146. Maier, N. R. F. Attention and inattention in rats. *J. genet. Psychol.*, 1930, **38**, 288–306.
147. Maier, N. R. F. The specific process constituting the learning function. *Psychol. Rev.*, 1939, **46**, 241–252.
148. Maier, N. R. F. *Frustration.* New York: McGraw-Hill, 1949.
149. Maier, N. R. F., & Schneirla, T. C. Mechanisms in conditioning. *Psychol. Rev.*, 1942, **49**, 117–134.
150. Maller, J. B. Personality tests. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 170–213.
151. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 431–465.
152. Montgomery, K. C. An experimental study of reactive inhibition and conditioned inhibition. *J. exp. Psychol.*, 1951, **41**, 39–51.
153. Montgomery, K. C. "Spontaneous alternation" as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, **42**, 82–93.
154. Montgomery, K. C. The relation between exploratory behavior and spontaneous alternation in the white rat. *J. comp. physiol. Psychol.*, 1951, **44**, 582–589.
155. Montgomery, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, **45**, 50–57.
156. Montgomery, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287–293.
157. Mosier, C. I. Psychophysics and mental test theory. II. The constant process. *Psychol. Rev.*, 1941, **48**, 235–249.
158. Mowrer, O. H. *Learning theory and per-*

sonality dynamics. New York: Ronald Press, 1950.

159. Mowrer, O. H., & Jones, Helen M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369–386.

160. Muenzinger, K. F. Plasticity and mechanization of the problem box habit in guinea pigs. *J. comp. Psychol.*, 1928, **8**, 45–70.

161. Muenzinger, K. F., Koerner, L., & Irey, E. Variability of an habitual movement in guinea pigs. *J. comp. Psychol.*, 1929, **9**, 425–436.

162. Muenzinger, K. F., & Wood, A. Motivation in learning. IV. The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *J. comp. Psychol.*, 1935, **20**, 95–106.

163. Neprash, J. D. Reliability of questions in the Thurstone Personality Schedule. *J. soc. Psychol.*, 1936, **7**, 239–244.

164. O'Kelly, L. I. An experimental study of regression. I. Behavioral characteristics of the regressive response. *J. comp. Psychol.*, 1940, **30**, 41–53.

165. Pascal, G. R., & Suttell, B. J. *The Bender-Gestalt Test*. New York: Grune & Stratton, 1951.

166. Pauli, R. Beiträge zur Kenntnis der Arbeitskurve. *Arch. ges. Psychol.*, 1936, **97**, 465–532.

167. Pauli, R. Die Arbeitskurve als ganzheitlicher Prufungsversuch (als Universaltest). *Arch. ges. Psychol.*, 1938, **100**, 401–423.

168. Paulsen, G. B. The reliability and consistency of individual differences in motor control. I. *J. appl. Psychol.*, 1935, **19**, 29–42.

169. Paulsen, G. B. The reliability and consistency of individual differences in motor control. II. *J. appl. Psychol.*, 1935, **19**, 166–179.

170. Persky, H. Response to a life stress: evaluation of some biochemical indices. *J. appl. Physiol.*, 1953, **6**, 369–374.

171. Philip, B. R. Studies in high speed continuous work. I. Periodicity. *J. exp Psychol.*, 1939, **24**, 499–509.

172. Philpott, S. J. F. Fluctuations in human output. *Brit. J. Psychol. Monogr. Suppl.*, 1933, **6**, No. 17.

173. Pintner, R., & Forlano, G. Four retests of a personality inventory. *J. educ. Psychol.*, 1938, **29**, 93–100.

174. Rabin, A. I. Fluctuations in the mental level of schizophrenic patients. *Psychiat. Quart.*, 1944, **18**, 78–81.

175. Rabin, A. I. Test constancy and variation in the mentally ill. *J. gen. Psychol.*, 1944, **31**, 231–239.

176. Rapaport, D. *Emotions and memory*. New York: International Universities Press, 1950.

177. Rashevsky, N. *Advances and applications of mathematical biology*. Chicago: Univer. of Chicago Press, 1940.

178. Reymert, M. L. The personal equation in motor capacities. *Scandin. Sci. Rev.*, 1923, **2**, 177–222.

179. Richardson, L. F. Dr. S. J. F. Philpott's wave theory. *Brit. J. Psychol.*, 1952, **43**, 169–176.

180. Riley, D. A., & Shapiro, A. N. Alternation behavior as a function of effortfulness of task and distribution of trials. *J. comp. physiol. Psychol.*, 1952, **45**, 468–475.

181. Rimoldi, H. J. A. Personal tempo. *J. abnorm. soc. Psychol.*, 1951, **46**, 283–303.

182. Robinson, E. S., & Bills, A. G. Two factors in the work decrement. *J. exp. Psychol.*, 1926, **9**, 415–443.

183. Roseman, M. Differential flexibility in normal and neuropsychiatric subjects on a test of closure behavior. Paper read at meeting of South. Soc. for Phil. and Psychol., Roanoke, Va., 1951.

184. Rosenzweig, S., & Mirmow, E. L. The validation of trends in the children's form of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 306–314.

185. Roshal, J. M. G. Changes in behavior variability with psychotherapy. *Amer. Psychologist*, 1952, **7**, 343. (Abstract)

186. Rothkopf, E. Z., & Zeaman, D. Some stimulus controls of alternation behavior. *J. Psychol.*, 1952, **34**, 235–255.

187. Sanders, M. J. An experimental demonstration of regression in the rat. *J. exp. Psychol.*, 1937, **21**, 493–510.

188. Sarvis, B. C. An experimental study of rhythms. *Psychol. Monogr.*, 1933, **44**, No. 1 (Whole No. 197).

189. Saudek, R. *The psychology of handwriting*. New York: George H. Doran Co., 1926.

190. Schmidtke, H. Uber die Messung der psychischen Ermudung mit Hilfe des Flimmertests. *Psychol. Forsch.*, 1951, **23**, 409–463.

191. Schofield, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychol.*

Monogr., 1950, **64**, No. 5 (Whole No. 311).

192. Schofield, W. A further study of the effects of therapies upon MMPI responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 67–77.

193. Seashore, C. E., & Kent, G. H. Periodicity and progressive change in continuous mental work. *Psychol. Rev. Monogr. Suppl.*, 1905, **6**, 47–101.

194. Senders, V. L., & Sowards, A. Analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1952, **65**, 358–374.

195. Shakow, D., & Huston, P. E. Studies of motor function in schizophrenia. I. Speed of tapping. *J. gen. Psychol.*, 1936, **15**, 63–106.

196. Siipola, Elsa, Kuhns, Florence, & Taylor, Vivian. Measurement of the individual's reaction to color in ink blots. *J. Pers.*, 1950, **19**, 153–171.

197. Skinner, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, **30**, 495–503.

198. Smith, B. B. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. *Brit. J. Psychol.*, 1951, **42**, 250–268.

199. Smith, G. J., & Klein, G. S. Cognitive controls in serial behavior. *J. Pers.*, 1953, **22**, 188–213.

200. Solomon, R. L. The influence of work on behavior. Unpublished doctor's dissertation, Brown Univer., 1947.

201. Solomon, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1–40.

202. Solomon, R. L. A note on the alternation of guesses. *J. exp. Psychol.*, 1949, **39**, 322–326.

203. Spearman, C. *The abilities of man.* New York: Macmillan, 1927.

204. Stalnick, S. A study of correlation between self-consistency and positive attitudes toward self. Unpublished master's thesis, Univer. of Chicago, 1953.

205. Susukita, T. Eine Untersuchung uber die Struktur des Charakters des japanischen Schulkindes. *Tohoku Psychol. Folia*, 1942, **9**, 1–42.

206. Taylor, D. M. Consistency of the self-concept. Unpublished doctor's dissertation, Vanderbilt Univer., 1953.

207. Taylor, J. G. Behavior oscillation and the growth of preference. *Psychol. Rev.*, 1949, **56**, 77–87.

208. Taylor, J. G. Reaction latency as a function of reaction potential and behavior oscillation. *Psychol. Rev.*, 1950, **57**, 375–389.

209. Telford, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1–36.

210. Thorndike, E. L. *The fundamentals of learning.* New York: Teachers College, Columbia Univer., 1916.

211. Thorndike, E. L. The variability of an individual in repetitions of the same task. *J. exp. Psychol.*, 1923, **6**, 161–167.

212. Thorndike, E. L. The refractory period in associative processes. *Psychol. Rev.*, 1927, **34**, 234–236.

213. Thorndike, E. L. *Human learning.* New York: Appleton-Century-Crofts, 1931.

214. Thouless, R. H. Test unreliability and function fluctuation. *Brit. J. Psychol.*, 1936, **26**, 325–343.

215. Thurstone, L. L. Psychophysical methods. In T. G. Andrews (Ed.), *Methods of psychology.* New York: Wiley, 1948. Pp. 124–157.

216. Tolman, E. C. Purpose and cognition: the determiners of animal learning. *Psychol. Rev.*, 1925, **32**, 285–297.

217. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

218. Turner, W. D. Intra-serial effects with lifted weights. *Amer. J. Psychol.*, 1931, **43**, 1–25.

219. Tussing, L. Perceptual fluctuations of illusions as a possible physical fatigue index. *J. exp. Psychol.*, 1941, **29**, 85–88.

220. Underwood, B. J. Learning. *Ann. Rev. Psychol.*, 1953, **4**, 31–58.

221. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44**, 273–282.

222. Walton, R. D. Relations between amplitude of oscillations in short period efficiency and steadiness of character. *Brit. J. Psychol.*, 1936, **27**, 181–188.

223. Webb, W. B., & De Haan, H. Wechsler-Bellevue split-half reliabilities in normals and schizophrenics. *J. consult. Psychol.*, 1951, **15**, 68–71.

224. Weber, C. O. Function-fluctuation and personality trends of normal subjects. *Amer. J. Psychol.*, 1938, **51**, 702–708.

225. Weitz, J., & Wakeman, M. L. "Spon-

taneous" alternation and the conditioned response. *J. comp. Psychol.*, 1941, **32**, 551–562.

226. Wertheimer, M. An investigation of the "randomness" of threshold measurements. *J. exp. Psychol.*, 1953, **45**, 294–303.

227. Wingfield, R. C., & Dennis, W. The dependence of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135–147.

228. Woodrow, H. Quotidian variability. *Psychol. Rev.*, 1932, **39**, 245–256.

229. Wulfeck, W. H. Motor function in the mentally disordered. III. Intra-individual consistency of expressive movement in psychotics, psychoneurotics, and normals. *J. Psychol.*, 1941, **11**, 151–160.

230. Yoshioka, J. G. An alternation habit in rats in a simple maze. *J. genet. Psychol.*, 1929, **36**, 257–266.

231. Yoshioka, J. G. Weber's law and the discrimination of maze distance in the white rat. *Univer. Calif. Publ. Psychol.*, 1929, **4**, 155–184.

232. Zeaman, D., & House, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177–186.

233. Zubin, J. A technique for measuring like-mindedness. *J. abnorm. soc. Psychol.*, 1938, **33**, 508–516.

Received June 10, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 3, 1955

SPECIAL REVIEW

SOME RECENT BOOKS ON COUNSELING AND ADJUSTMENT

EDWARD JOSEPH SHOBEN, JR.
Teachers College, Columbia University

ARBUCKLE, D. S. *Student personnel services in higher education.* New York: McGraw-Hill, 1953. Pp. x+352. $4.50.

ARSENIAN, S., & MCKENZIE, F. W. *Counseling in the YMCA.* New York: Association Press, 1954. Pp. 126. $2.00.

CRAIG, R. C. *The transfer value of guided learning.* New York: Bureau of Publications, Teachers College, 1953. Pp. viii+85. $2.75.

DRIVER, HELEN IRENE. *Multiple counseling.* Madison, Wis.: Monona Publications, 1954. Pp. 280. $5.00.

GRIFFITHS, W. *Behavior difficulties of children as perceived and judged by parents, teachers, and children themselves.* Minneapolis: Univer. of Minnesota Press, 1952. Pp. xii +116. $3.00.

HATHAWAY, S. R., & MONACHESI, E. D. *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. of Minnesota Press, 1953. Pp. vi+153. $3.50.

HOYLES, J. A. *The treatment of the young delinquent.* New York: Philosophical Library, 1952. Pp. viii+271. $4.75.

HUMPHREYS, J. A., & TRAXLER, A. E. *Guidance services.* Chicago: Science Research Associates, 1954. Pp. xvii+438. $4.75.

KATZ, B., & LEHNER, G. F. J. *Mental hygiene in modern living.* New York: Ronald Press, 1953. Pp. x+544. $4.50.

KNAPP, R. H. *Practical guidance methods.* New York: McGraw-Hill, 1953. Pp. xi+320. $4.25.

LINDGREN, H. C. *Psychology of personal and social adjustment.* New York: American Book Co., 1953. Pp. ix+481. $4.50.

LITTLE, W., & CHAPMAN, A. L. *Developmental guidance in secondary school.* New York: McGraw-Hill, 1953. Pp. xi+324. $4.50.

PEPINSKY, H. B., & PEPINSKY, PAULINE NICHOLS. *Counseling: Theory and practice.* New York: Ronald Press, 1954. Pp. vii+307. $4.50.

RECKTENWALD, L. N. *Guidance and counseling.* Washington, D. C.: Catholic Univer. of America Press, 1953. Pp. xiii+192. $3.25 ($2.50 paper cover).

SANDERSON, H. *Basic concepts in vocational guidance.* New York: McGraw-Hill, 1954. Pp. xiii+338. $4.50.

SHAW, F. J., & ORT, R. S. *Personal adjustment in the American culture.* New York: Harper, 1953. Pp. ix +388. $4.00.

TYLER, LEONA E. *The work of the counselor.* New York: Appleton-Century-Crofts, 1953. Pp. xi+323. $3.00.

WARTERS, JANE. *Techniques of counseling.* New York: McGraw-Hill, 1954. Pp. viii+384. $4.75.

Perhaps the most telling sociological phenomenon in psychology is its rapid professionalization during the decade since the end of World War II. In a scant ten years, an essentially academic discipline has transformed large segments of itself into a form of public service, meeting in the process the problems with which a service profession must cope: relationships with other professions, the formulation of ethical codes, provisions for internal policing of professional practice, the development of training standards and training opportunities, concerns about legislative recognition, the maintenance of public relations, and many others.

One problem that cuts deeper than others, however, is seldom faced squarely. Psychology's professionalization has occurred largely as a response to a social demand intensified and given justification, although not created, by wartime experience. Such a demand takes primary form as a market for new workers to deal helpfully with the troubles of unhappy people. Since the knowledge necessary to this task has not yet been fully discovered and developed, a need is created for a kind of substitute for knowledge, a literature of practical wisdom and summarized clinical experience. And because the human problems toward which professional psychology is directed are genuinely important and poignant, because something crucial in the lives of his clients is at stake, the professional psychologist himself often is motivated to seize desperately on any technique or idea that has the appearance of usefulness. When the chips are down, as they generally are in professional practice, skepticism about one's own resources is a luxury that few can afford.

One outcome of this state of affairs is the current rift between "pure" and "applied" psychologists. Possibly a bit jealous of the social prominence and income of their colleagues in clinical and counseling positions, research men and serious theorists are likely to judge the literature and oral pronouncements in these fields, often with some justification, as softheaded and close to meaningless. On the other hand, made defensive by such charges and harried enough by their applied responsibilities, the professionals tend to reply in hasty irritation that the work of the scientists is unfeeling and irrelevant.

Actually, of course, social need has always outrun available knowledge, generally providing the essential spur to its pursuit. To say that professional practice must rely on hunch and accumulations of uncontrolled experience is to say nothing derogatory *so long as one knows what is happening*. The significance of the present split in psychology is that the element of faction makes it harder for people to acknowledge the difference between knowledge and practical approximations, to work in concert for the replacement of the latter by established fact and soundly based theory, and to formulate from professional practice fruitful hypotheses for the advancement of psychological science.

The professionalization of psychology, then, provides not only opportunities for the discipline's being of service to a needful public but also opportunities for the enrichment of psychological science through the hypothesis-forming potential of applied work. Its danger lies in a splitting of science and profession to the detriment of both. The extent to which this divisive tendency has permeated or been controlled in the field of counseling psychology and adjustment can be estimated from these 18 recent books.

CONCEPTS OF ADJUSTMENT

Willy-nilly, what counselors do in their relationships with clients is determined in part by their implicit or explicit notions of what constitutes effective personal adjustment and their unformulated or articulate concepts of how it is attained. Most counselors, clinicians, and social psychologists find themselves involved in difficult and ongoing struggles to clarify such questions as the relationship of social conformity to the gratification of individual needs, of articulation with the cultural group to personal spontaneity and novelty, and of the acceptance of social regulation to the promptings of one's individual experience conceptualized as conscience.

Shaw and Ort treat these problems courageously and in a sophisticated fashion in a book that blends effectively clinical experience with research evidence and theories that have had some brush with the laboratory as a proving ground. They argue that the individual's adjustments are best understood as ways of interacting with other people and as products of primarily social experience. This emphasis on one's history and current behavior in interpersonal, social, and cultural contexts is formulated within a conceptual framework of reinforcement learning theory in which personality development and functioning is comprehended as the result of rewards and punishments experienced in the social learning environment.

On the other hand, they are most mindful of the problem of adequacy in personal adjustment and wrestle manfully with the criterion of mere social conformity to which a sociological point of departure often leads. Their resolution of the difficulty is hardly a final one. Shaw and Ort rightly hold that the ability to act in accordance with social standards is an essential part of effective interaction with one's inescapable interpersonal world but that it must be balanced by something they call "integration." This factor is defined as the maintenance of oneself in one's environment, and behavior disorders are conceived as deviations from adequate integration.

This concept involves two interlocking difficulties. The first is the theoretical embarrassment that derives from the attempt to fit the notion of self-maintenance, deriving from a phenomenological tradition, into the objectivist behavior theory that is the conceptual groundwork for *Personal Adjustment in the American Culture.* This task of unifying different psychological theories is a most worthy one, but the attempt here results in a weakening of the rigor of reinforcement theory and a loss of the intuitive impact that seems to have made phenomenology attractive for many applied workers. Second, the effort to deal with integration apart from conformity lands the authors in the problem of accounting for integrative behavior that runs *counter* to their social experience, of developing explicit criteria for judging when nonconformity represents adequate adjustment and when it suggests mere rebellion or lack of social awareness, and of providing a *psychological* basis for understanding integrative adjustments that lie outside cultural norms.

Two comments seem in order. The first is that one of the sources of uncertainty and a kind of fuzziness of thought in much of applied psychology may be the necessity of involvement in questions which are essentially ethical or philosophical in character. In a sense, Shaw and Ort accepted the task of defining "the good life" or "the elements of good conduct." A. E. Taylor in his life of Socrates (7) makes much of the Socratic "sense of the importance of implicit obedience to lawful authority" without falling into the "vice of exalting the mere letter of the law above its spirit." It would seem that Shaw and Ort's courageous book, for all its labors, says little more on this difficult topic of the balance between conformity and integration. There is a temptation to say that psychology should stick to its traditional last, but as May (5) and Mowrer (6) point out, the question of what kind of behavior is associated with human happiness is a legitimate one for science to ask and an inevitable one for the psychological professions.

The second point to be made is that few books illustrate as well as Shaw and Ort's how far a knowledgeable fusion of experimentally based theory and clinically based hunch can achieve two goals. One is the clarification of processes observed in the consulting room and in the field. Defensive operations, the phenomena of age grading and other forms of role taking, masculinity and femininity, and occupational strivings become less complex and communicable in more manageable terms when analyzed according to a theoretical system the strengths and weaknesses of which are assessable in part through laboratory tests. The other is the enrichment of a theoretical structure by seeing how well it can be made to cover practical observations. Even though it was written as a relatively low-level textbook, this volume will repay close reading by those interested in developing research problems in that important area that lies between restricted theory and loose practice.

The volume by Katz and Lehner, on the other hand, is theoretically much less ambitious. Its pretensions lie in the direction of explaining human behavior to unsophisticated readers *and* of demonstrating to these readers how they may achieve happier adjustments in their own lives. As such, *Mental Hygiene in Modern Living* is a kind of cross between text and self-help book. This hybridization could conceivably make for certain strengths. For example, the encyclopedic coverage virtually assures most readers of finding something useful in its pages. Like an encyclopedia, however, the book lacks system and thoroughly confuses practical wisdom with verified knowledge and substantiated theory. Nowhere does it come to grips with the problem of what constitutes *positive* adjustment in any cogent way, and the discussion of maladaptive behavior, including the psychoses, leads one to ponder the observation that clinicians and counselors so often seem more at home with psychological ills than with psychological health. Even the technical vocabulary seems wanting in words to describe effective adjustment and the conditions which determine it. It is little wonder, although a trifle alarming, that so little research is done and so little theorizing devoted to the development of personal assets when professional attention is directed so strongly to pathology. One unfortunate result is that discussions of positive personality growth have, as they do in this book, the leaden ring of clichés sounding through them. Saddest of all, this fault cannot be charged directly to the authors. Citing a wide literature, they refer to virtually no studies directly concerned with psychological health; one guesses that they found none. For counseling psychology in particular, with its official emphasis (1) on the positive and the "normal," this absence of careful thought and empirical evidence on the problem of what constitutes normality and its determinants seems to be a deadly realm of ignorance.

Lindgren makes some attempt to deal with this worrisome gap in knowledge and theory in a readable book that pays little attention to the major forms of psychopathology and concentrates on the understanding of "ordinary people." His approach is not discernibly systematic, like Shaw and Ort's, but neither is it a wide-ranging mass of unconnected propositions and unorganized bits of information, like Katz and Lehner's. Refreshingly, Lindgren recognizes human resiliency, the capacity to absorb considerable amounts of frustration without serious damage, as a primary factor in mental health; and he also deserves much credit for bringing *thinking* back to the list of psychological functions of importance to counselors. Dynamic psychology, in its reaction to the rationalist tradition and its proper underscoring of affective and conative determinants of behavior, has forgotten until very recently Freud's (4) remark, "The voice of the intellect is a soft one, but it does not rest until it has gained a hearing." It is to Lindgren's credit that he gives due emphasis to the role of thought in solving common human difficulties without falling back on the notion that the problems of people are always at bottom intellective affairs.

Nevertheless, the emphasis on pathology that has characterized modern psychology, especially in its professional aspects, seems to provide the undercurrent for this, as for other, volumes. The conception of man implicit here is one that depicts him as being buffeted by his circumstances, protected only by his degree of native toughness and his wisdom in finding "therapeutic" experiences either within or without the consulting room. Everyday social life, work, and religion (conceived naturalistically as a form of ready communication with a group and its leader) are all discussed as forms of therapy, i.e., as *corrective.* This view may be quite correct, but it is not unchallengeable. One can legitimately raise the question of whether work and personal relationships cannot have positive effects in themselves, not as correctives to unavoidable ills, just as one can ask if the inescapable frustrations and compromises of daily living necessarily require therapeutic relief in mature people. This rather different idea of human functioning might lead to a re-evaluation of experiment and theory and the development of a fruitful new line of psychological research on the nature and antecedents of normality.

One might expect that these concerns for the *nature* of adequate adjustments, involving something very close to ethical and philosophical considerations, might find more explicit attention in two books written from religious points of view. Hoyles' book on delinquency is by a minister of the Church of England, and Recktenwald's discussion of guidance procedures is by a Roman Catholic. The first is a humane and intelligent argument for rehabilitative rather than punitive treatment of juvenile offenders. The second is a workmanlike and comprehensive brief overview of guidance techniques in school settings. Neither contains anything particularly novel, and Hoyles seems more preoccupied with presenting a challenge to the Anglican Church than with analyzing the character of criminal adjustments among young people. The important thing about them which justifies considering them together here is their mutual lack of attention to the problem of what constitutes adequate or integrative behavior. Except for a very few random pages, Recktenwald's book could have been written by a non-Catholic, and Hoyles uses the language of moral theology to say essentially the same things that anybody would say who conceives of delinquency as a form of behavioral pathology rather than a kind of wilful violation of social tenets. Neither grapples as frankly as do Shaw and Ort, for example, with the problem of adequate adjustment or the behavior patterns associated with "the good life."

RESEARCH STUDIES

If the nature of integrative adjustment is given little attention in these books on mental hygiene, it is not surprising to find a similar lack of concern in the three research monographs by Hathaway and Monachesi, Craig, and Griffiths. The noteworthy thing about these three very different publications is that they represent the vigor with which sound empirical work is being pursued in a professional field. All three, however, have only remote connections with comprehensive theory, which may reflect again the lack of articulation between the scientific and the systematic tradition and the professional one in psychology generally. Their diversity is also worth noting as indicative of the number of problems accepted as the investigative responsibility of counseling psychologists and those concerned with school guidance.

The Hathaway and Monachesi monograph is an exceptional piece of work for two reasons quite apart from the high level of research sophistication that marks it. First, concerned with the prediction of delinquent behavior in children, it avoids the pitfalls of retrospective data. Instead of studying the attributes of those already guilty of delinquent acts as opposed to those of youngsters who are not delinquent, Hathaway and his co-workers began with the examination of large numbers of subjects below the age at which delinquency tends to occur. Predictions were made at that point, and their accuracy was determined by a close follow-up of the children.

Second, this little book is a splendid exhibition of what can be done with structured personality tests. Using the MMPI, these researchers concerned themselves with profile patterns rather than individual scores, thus introducing a regard for the complexity of test responses similar to that insisted upon by Rorschach workers and the users of other projective techniques. The method of coding profile patterns, however, does not sacrifice the rigor and exactness associated with quantitative scores but usually purchased at the cost of a good deal of information about the respondents. In short, the methodological implications of this study are that objective personality scales can be used in such a way as to retain precision without exorbitant cost in what Cronbach (**2**) has called "bandwidth." In a field where assessment procedures and predictive devices often yield blurred and fuzzy results, the model provided here assumes particular importance.

As for findings, Hathaway and Monachesi present convincing data to indicate that adult test patterns of the amoral psychopath and the hypomanic individual tend to occur with predictively significant frequency among predelinquents, whereas the occurrence of test patterns characteristic of adult neurosis seem to exercise an inhibiting influence on the development of delinquent behavior. Similarly, those test patterns which show no high deviations and are generally considered as "normal" are remarkably predictive of nondelinquent adjustments. This study strongly suggests that the interpretation of delinquency as a form of neurotic behavior overlooks fundamental differences in neurotic and criminal adjustments and implies that the dynamics of delinquency are not to be found in repressions and self-derogating tendencies.

On the other hand, there is little in this monograph to shed much light on the antecedents of delinquency or its modification. Meaningful investigations of the experiential determinants of delinquent behavior, of the relevant learning environments in which delinquent patterns are acquired, are yet to come. Until such inquiries are undertaken with the same vigor, sophistication, and insistence on exactness that one finds in this prognostic study, it is unlikely that a useful theory of delinquent adjustment will be developed to replace current clichés, hunches, or scattered bits of knowledge.

Griffiths' study of children's behavior difficulties is not concerned with delinquency. Rather, he was concerned with the kinds of behavior problems that children themselves think they have. Studying youngsters from six to fourteen, he finds that awareness and acceptance of social rules and regulations increase with age, that younger children tend to be aware of overt aggression as a source of trouble but only later to develop a sense of uneasiness about submissive and withdrawing tendencies in themselves, that children generally are surprisingly aware of how both parents and teachers would like them to change their behavior, and that children from the middle socioeconomic group are more conforming to adult norms than those from other classes.

The most outstanding thing about this study is its documentation of the effects of socialization. As age—which means social experience—increases, children become increasingly sensitive to adult demands and norms of conduct and even at relatively early levels show surprising insight into adult desires to have them change their behavior in certain identifiable ways. Moreover, children seem to discriminate fairly early the different demands made of them by different adults. Fathers, for example, are perceived as persons who primarily ask that they not be disturbed. Teachers are people predominantly interested in an orderly and efficient classroom, although youngsters seem a bit more aware of the interest of teachers (in contrast to parents) in the modification of withdrawing and submissive traits. Mothers, like teachers, want smooth-running homes and evaluate children's behavior accordingly.

Methodologically simple, this monograph provides a wealth of information about children's evaluations of their own behavior and relevant to an increased understanding of the socialization process. Again, however, the findings are not related to any generative kind of theory that would lead readily to more systematic research on the socialization process or the antecedents of children's attitudes. The importance of this problem also requires comprehensive study to deal with such questions as the influence of perceived adult norms on peer relationships, the development of conscience, and the development of personality. For example, does the child who becomes aware of adult conduct norms at an early age tend to develop a greater degree of responsiveness to authority than the youngster who acquires this kind of awareness later? To what extent is this kind of awareness related to the type of rewards and punishments typical of the home and to the pattern of child-parent relationships generally? Griffiths' study provides an excellent point of departure for theoretically important investigations of the antecedents of all kinds of adjustment problems, including that of "adequate" or integrative adjustment patterns.

Craig's monograph on guided learning is in the tradition of transfer-of-training experiments of the Thorndikian type, but its implications are somewhat broader than most such studies. Working with recent college graduates as subjects and using a task consisting of multiple-choice verbal test items, Craig found that giving learners a short statement of principles common to a group of items reduced errors, increased the efficiency of solutions, and aided the process of discovering the basis for correct response to other items. This effect became more pronounced as a function of the difficulty of the learning situations. While conceived within narrow and well-controlled laboratory conditions, the experiment suggests that interpersonal as well as intellectual problem solution might well be facilitated by the development of principles common to many situations through participation in discussion groups, orientation classes, and individual guidance conferences. Growing out of the tradition of law-of-effect learning studies, Craig's inquiry has implications that amount to ready-made hypotheses for counseling psychologists to put to test. This kind of research would have the considerable advantage of being tied to a comprehensive theory of behavior that could be expanded in two directions: application in a knowledgeable kind of way to significant social problems and theoretical extension to affective and conative spheres.

COUNSELING TECHNIQUE

The great bulk of these books is devoted to technique, the methods to be employed by the practical counselor. This state of affairs is perhaps a function of many factors. One is the historical emphasis on practicality and tangible action in American culture, one current manifestation of which is the spate of how-to-do-it books in many areas. Another is the intensity of the demand out of which the professionalization of psychology has grown. With a greater awareness and acknowledgment of human problems, there is a greater insistence that preventive and remedial treatment be supplied by *somebody*. One outcome seems to be a crop of manuals and books of instructions for those who are pulled into this vacuum of need, often without the fundamental training upon which professional practice must rest.

The striking thing about the present collection of nine volumes is their similarity and degree of overlap in content. The books by Arbuckle, Humphreys and Traxler, Knapp, Little and Chapman, Recktenwald, and Warters all say essentially the same things in essentially the same ways. One marvels at a market that can absorb such comparable publications. It is not that these books are lacking in workmanship or competence; they have both qualities. But they are all very much in the same vein of describing activities by which a counselor can keep himself busy doing technical things.

Of these technical things, two are given major predominance. One is testing, the other is record keeping. Both, of course, are important, but one wonders if lists of psychometric instruments and lengthy discussions of cumulative records are the core of the counseling process. Counselors certainly need methods of sharpening the precision of their observations and ways of estimating how a given client or student compares with his fellows. Tests are very relevant. Likewise, counselors must know how to keep account of the development and progress of those to whom they are responsible. Record-keeping devices are very much in point. The present manuals, however, devote very little space to the ways in which these technical aids can be used to advance the aim of the counseling process, *the helpful modification of client behavior*. Neither is there much space afforded to the use of the data collected by such instruments in the advancement of knowledge. If counseling psychology is to remain a part of *psychology* and not to degenerate into mere technician status, then surely research and careful thought about human problems as they are encountered on campuses, in industry, and in the consulting room generally must be willingly assumed functions of practitioners of the counselor's craft.

If this point can be regarded as debatable, however, another can hardly be. The essence of the counselor's job is the modification of client behavior through face-to-face contacts. It is *counseling*. The proportion of pages in six fat volumes concerned directly with this face-to-face process of behavior modification is of the order of eight per cent! One suspects that under professional pressure it is somehow easier to be occupied with testing programs and the maintenance of records than with direct personal service. One also suspects that the lack of real knowledge about the counseling process is an overwhelming if inarticulate determinant of this strange fact. Is it possible that the how-to-do-it manuals are premature in many ways and that more research and more thought on the counseling process and its outcomes are prerequisites yet to be achieved for preparing truly meaningful books on technique? The profession is young, and its undergirding of basic knowledge may be too weak still to support such a structure of textbooks. The same comments apply to Arsenian's little book, a set of instructions for YMCA secretaries who find themselves involved in counseling relationships.

Two books are exceptions. One is Driver's discussion of the technique of multiple counseling, the other Tyler's sophisticated and informed *The Work of the Counselor*. The first is an extensive and interesting presentation of small-group discussion methods of modifying behavior drawn primarily from a fusion of client-centered psychotherapy and group dynamics. The book is noteworthy because of its extensive case excerpts, its serious if rather loose attempts to evaluate procedures and outcomes, and its step-by-step descriptions of multiple counseling groups and their development. Anyone interested in group counseling methods will find rewarding reading here, and the challenge to rationalize the method and to subject it to rigorous investigation is considerable. Research hypotheses abound in its pages, and the book's freedom from a doctrinaire tone invites the test of empirical inquiry.

Tyler's volume, by far the best of the "practical" books on the present list, frankly acknowledges the status of counseling as an "art," dependent more on the accumulation of shared clinical experience than on research evidence and tightly constructed theory. Nevertheless, she argues that the rapidity of improvement in counseling procedures must rely on the soundness of relevant investigation and that intuition can be made articulate through research. Her chapters, therefore, are organized in terms of what counselors appear to do and think in their practice but are supplemented with research summaries reviewing the relevant research literature and clarifying issues on which research may throw some light. Unlike other authors, Tyler's greatest emphasis is on the interview, and her understanding of the relationship between observation on the one hand and behavioral modification on the other is insightful and seminal while appropriately humble.

Oddly, this fine book pays virtually no attention to personality dynamics or the problems of clarifying the concept of adjustment. This is the more remarkable since Tyler's is the only one of these volumes that makes it explicitly clear that the profession of counseling cannot be successfully followed by one whose only equipment is a bag of tricks, a knowledge of available tests, and ways of keeping records. Much of what she says implies that counselors require more than anything else an ever-developing knowledge of learning, perception, motivation, and human development, and the failure of her book to concern itself with these substantive issues in some degree is surprising and disappointing. There is a limit, however, to which an author can be taken to task for not writing a different book, and Tyler's contribution is too informed and too well based on general psychological knowledge and in the traditions of psychological science to permit much legitimate fault finding.

THEORETICAL CONTRIBUTIONS

As a corrective to the how-to-do-it manual, professional psychology has two major theoretical alternatives. It can attempt to systematize its observations and to develop a formal body of propositions independently of the rest of psychological science, risking a fragmentation and divisive force that might well deprive the profession of the intellectual and scientific bases on which most mature professions rest. Or, it can formalize its observations in terms of general psychological theory, a process which demands the assumption that such processes as learning, motivation, and perception are essentially similar whether studied in the laboratory or in the counselor's office.

The little book by Pepinsky and Pepinsky is an outstanding example of the latter possibility. With even more cogency and creativeness than Dollard and Miller's *Personality and Psychotherapy* (**3**), this work illustrates the advantages of relating clinical experience to a formal and elegant theory derived from experiment and the theorist's study. The Pepinskys both apply and enlarge a relatively rigorous Hullian model by using it to account for anxiety and unproductive behavior as these concepts are employed descriptively by counseling practitioners. With a minimum of concern for specific techniques, *Counseling: Theory and Practice* demonstrates how a counselor can better understand and communicate with his clients if he is thoroughly in possession of a comprehensive theory that is well fortified by careful research than if he attempts to rely on a bag of technical tricks or the unrelated and sometimes inconsistent assertions that occasionally pass for theory among those whose training is deficient in general psychology.

Indeed, a major thesis of the book is that professional practitioners must work *both* as applied clinicians and as scientists, each role facilitating the other. Experience as a counselor provides observations that allow the inductive formulation of general laws. Experience as a scientist permits the testing of these laws and the determination of which are useful, which require modification, and which must be scrapped as unproductive of either new insights into the counseling process or new hypotheses for research. Practice unleavened by theoretically oriented research responsibility leads to a softness of principle and a blurring of theoretical precision that is likely to reduce counseling to a matter of *mystique* and uncommunicable intuitions. Theory-oriented research without familiarity with practical problems and clinical events risks a narrowing and a rarefying that make theory lose significance and, worse, lose comprehensiveness.

Sanderson's *Basic Concepts in Vocational Guidance* is a very different matter, more illustrative of the first alternative, the construction of theory from professional observations without regard to more general psychological propositions. It is possible to read this volume without being reminded that human beings learn and perceive and manipulate symbols in thought or social communication! And yet the book is a good one, infused with the attitudes of science if not its general psychological content and reflecting a careful and thoughtful evaluation of a wealth of literature and counseling experience. Contrary to the how-to-do-it books, this one rests on the proposition that counselors rely too heavily on "diagnosis" (testing and record keeping), and too little on "the helping process itself" (behavioral change through face-to-face contact). Sanderson shares with the Pepinskys, then, the belief that counseling practice is facilitated more by a coherent body of ideas by which the clinician can understand his clients and his procedures than by a clutch of mechanical methods.

Further, his conception of the counseling role is a broad one, taking fully into account the motives and affects that revolve around the status aspects of occupational adjustment and the variety of noneconomic rewards bound up with work. But this very breadth of conception leads to disappointment when the reader finds nothing about the general psychology of personality or the research evidence and tentative principles formulated from it that bear on personal development, the making of decisions, or the socially acquired drives that determine so much of human activity. The alternative of reasonably precise but discreet and unsystematized concepts drawn from counseling experience itself may help to clarify the counselor's present work, just as Sanderson's stress on understanding the helping process as contrasted with reliance on technique may help to increase the actual amount of service given to clients. It is most unlikely, however, to advance the field, to stimulate a search for new knowledge of relevance to the process or its outcomes, or to increase the understanding of those general human functions which are essentially the subject of the counselor's concern, just as they are the subject of interest of friends, experimenters, and virtually all others who find the study of man inevitable as well as fascinating.

Neither of these two competent books deals with the problem that seems so central even while it is so thoroughly ignored. What constitutes effective adjustment or integrative behavior? Both Sanderson and the Pepinskys are essentially silent on the question, although the latter skirmish with it briefly in a provocative chapter on the assessment of counseling processes and outcomes. It seems somehow unlikely that fruitful outcome-research can be accomplished until this basic element in establishing a criterion is thought through.

SOME GENERAL COMMENTS

If the present list of books is a fair sample of the present state of counseling psychology, there is room for worry over the effects of rapid professionalization on psychology as science. With the public demand for service high, there is apparently a considerable interest in publications dealing with *techniques*, less in books seriously occupied with *knowledge and understanding*. There seems to be a tendency for practice to divorce itself from science that could be detrimental to both. Should counselors become only technicians and general psychologists only runners of rats and manipulators of brass instruments, both psychology and a needful public will suffer. The professionalization of large elements in psychology is welcome not only because it provides an opportunity to serve troubled people but because it opens new vistas for psychological research and the understanding of behavior. The problems are large enough so that their solutions probably lie in the collaboration of practitioner and rat man or tachistoscope operator. Without such collaboration, productive solutions seem far distant.

But there is also evidence to justify ending this review on other than a Cassandra note. The books by the Pepinskys and Shaw and Ort and the three research monographs all indicate an investigative vigor and a retained relationship with general psychology that holds much in the way of theoretical promise, just as Tyler's volume and, to a lesser degree, Sanderson's suggest that careful thought, a proper appreciation of research, and a knowledge of available evidence from the laboratory and the field still have a stout influence on practice. Division is by no means inevitable, and perhaps only a little awareness of the issues is necessary to insure the development of a useful profession firmly rooted in psychological knowledge and rigorously developed theory.

REFERENCES

1. AMERICAN PSYCHOLOGICAL ASSOCIATION, DIVISION OF COUNSELING AND GUIDANCE, COMMITTEE ON COUNSELOR TRAINING. Recommended standards for training counseling psychologists at the doctorate level. *Amer. Psychologist*, 1952, **7**, 175–181.
2. CRONBACH, L. J. Report on a psychometric mission to Clinicia. *Psychometrika*, 1954, **19**, 263–270.
3. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
4. FREUD, S. *The future of an illusion.* New York: Liveright, 1928.
5. MAY, R. *Man's search for himself.* New York: Norton, 1953.
6. MOWRER, O. H. Neurosis and psychotherapy as interpersonal processes: a synopsis. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 69–94.
7. TAYLOR, A. E. *Socrates.* Garden City, N. Y.: Doubleday, 1953.

CARMICHAEL, LEONARD. (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954. Pp. ix+1295. $12.00

The second edition of the *Manual* presents an excellent picture of child psychology. In this respect it is a better book than the first edition, which in the reviewer's opinion was out of date when it was issued.[1] This does not mean that this is a definitive work. It is, rather, an expansive, factious book. But it would be surprising if it were otherwise, considering the unfinished state of the frontier area with which it deals. At the same time it is an ambitious book that faithfully reflects the vigor and aspirations of this youthful science. The title of the volume, itself, expresses the ambitious tone of much of the book. A manual, by dictionary definition and by common usage, is a book of reference which can be conveniently carried in the hand. But this book's four and one-half pounds make it too big to be easily handled, and it is a book to be studied and argued with, not one to be referred to on the run for authoritative answers to standard problems.

The book is similar in organization and authorship to the first edition, seventeen of its nineteen chapters having the same titles and sixteen the same authors as the first edition (two chapters have different junior authors). Chapter changes from the first to the second editions are: "The Feebleminded Child" (Doll) is incorporated into a more inclusive chapter called "Psychopathology of Childhood" (Benda); "Maturation of Behavior" (McGraw) is omitted; "Social Development" (H. H. and G. L. Anderson) is added; "Adolescence" is written by Horrocks in the place of Dennis. A number of chapters in the new edition reprint large sections from the previous one in unchanged form. To be exact, 70 per cent or more of the following chapters are reprintings of the first edition: "The Onset and Early Development of Behavior" (Carmichael), "Animal Infancy" (Cruikshank), "The Neonate" (Pratt), "Physical Growth" (Thompson), "The Ontogenesis of Infant Behavior" (Gesell), "Learning in Children" (Munn), "Mental Growth" (Goodenough), "Research on Primitive Children" (Mead), "Behavior and Development as a Function of the Total Situation" (Lewin, with a supplement by Escalona), and "Gifted Children" (Cox). About half of this edition of the *Manual* is a reprinting of the first edition, and about half involves major revisions or new topics.

The ten chapters listed above will not be considered in detail here. The reader is referred to the earlier review for comments on them, as their analyses of the topics under consideration are essentially unchanged. These are research areas which the authors see as having been on a plateau in recent years so far as new developments are concerned. This does not mean, of course, that important additions, replications, and verifications have not occurred. For example, a keystone in the study of gifted children was put in place during this period by Terman and Oden's twenty-five year report.

The reviewer must repeat his objection to the chapter on the physical growth of children. He agrees completely with the author that "child behavior cannot be understood apart from . . . the physical body through which it perceives, reacts, and functions, and to which others react," (p. 292); but unfortunately he finds here no consideration of this problem. Can't we have a chapter on the *psychology* of physical differences in children in the next edition such as we have on the psychology of sex differences in this one?

---

[1] Roger G. Barker. Manual of child psychology, a special review. *Psychol. Bull.*, 1947, **44**, 162–170.

The chapters of the *Manual* where major changes appear, and the new chapters, may be individually mentioned as follows:

"Methods of Child Psychology" (John Anderson). Anderson wrote chapters with this title in the 1933 edition of Murchison's *Handbook* and the 1946 edition of the *Manual*. These, with the present essay, constitute a revealing exhibit of the rewards of two decades of preoccupation by psychologists with the problems of methodology. Anderson wrote approximately 11,000 words in 1933, 22,000 in 1946, and 33,000 in 1954 and while the increase in level may not be quite so marked as the increase in quantity, the reviewer perceives it as being nearly so. However, the scope of the chapter is too wide; it is, in fact, a survey of general psychological methodology. With a few changes in nomenclature and examples this essay would be as appropriate in a social or an experimental psychology as it is here. One hopes that the time will come when the general field will have been so well covered by others that Anderson can devote his attention to *special* problems of methodology in psychological research with children.

"The Environment and Mental Development" (H. E. Jones). Jones' review shows that there has been an active and profitable period of work on this area. He reports a welcome movement away from global nature-nurture studies to increased specification of the effects of particular environmental influences. Jones appears to take the position, without elaborating and emphasizing it sufficiently, however, that this movement must go still further. He says in connection with the Iowa studies "our chief need in this field is not for statistical methods of greater power and subtility but for more vigorous experimental procedures . . . " (681). If he means here, as the reviewer thinks he does, the need for better definition and control of environmental variables, he has raised a crucial point. In its context, Jones' statement seems to imply that, for example, it is not enough to study the effects of nursery school attendance upon intellectual development; it is necessary to define the significance of nursery school for children in psychologically meaningful concepts: perhaps in such variables as the stimulation, the attractiveness, the freedom, the rewards, and the learning opportunities of the nursery school situation. This points to an important requirement which Jones' review does little to meet. Although his critical evaluations of the technical adequacy of individual studies are excellent, he does little to organize and conceptualize the kinds of problems in this field. All "environmental" influences, ranging from culture, economics, social class, foster homes, hospitalization, and schooling to birth order, season of the year, nutrition, health, physical size, physiological maturity, and race are treated on the same level. Surely the time will soon come when some discrimination and sorting of problems within this field will be possible in terms of primitive theories of environmental variables.

"Character Development in Children—An Objective Approach (Vernon Jones). The author reports many new investigations of relevance to this topic, but the area of understanding appears to have been increased very little in the time since the first edition. It is clear that this is an area where the multiplying of facts is proving of little benefit. New directions are needed.

"Language Development in Children" (McCarthy). This is the longest chapter in the *Manual* (120 pages) with the largest number of references (776). It has been reorganized and greatly expanded for this edition. Extensive new sections cover the vocalization of infants and language disorders in relation to personality development. The problems considered cover a tremendous range including, for example, fetal sounds and the birth cry; vowel and consonant sounds in infant vocalization; vocabulary tests; amount and rate of talking; the function of language in the child's life; the effects of institutionalization and multiple birth upon language development; the correlation between language and motor, intellectual, and social development; language disorder syndromes and personality development; delayed speech; articulatory defects, stuttering. Despite McCarthy's very effective organization and condensation (including two four-page summarizing tables), the amount of material is so overwhelming that one leaves this chapter with a great weariness. Perhaps the basic difficulty is that "spoken language" is in itself not a useful category of behavior; perhaps it is only phenotypic, symptomatic behavior like hand behavior and foot behavior. In any case, to make future reviews more effective, some means of focusing upon the psychologically most meaningful and timely issues is essential. As it stands, this chapter provides a useful guide to a vast literature.

"Psychological Sex Differences" (Terman and Tyler). This chapter has been largely rewritten for this edition of the *Manual*, and it might well serve as a model for the kind of exposition suited to this book: the problem of the chapter is clear, the evidence is not beyond the limits of what can be handled in the space available, the authors make explicit the limits of their review and of the bibliographical sources covered, they include only those publications most relevant to the trend of the evidence as they see it. This method reveals clearly the status of the problems as viewed by the authors. Developments in the interval since the first edition of the *Manual* are more adequate data which confirm earlier findings and which reinforce the authors' earlier caution regarding possible biological and cultural roots of psychological sex differences. Theories and speculations about the possible lines of influence of these factors are not covered. This is a feature of the problem which some might consider of value.

"Emotional Development" (Jersild). This is a completely reorganized and greatly expanded survey of emotionality in children. Jersild places his discussion within an explicitly stated conception of emotion. He considers not only research findings, but various theories and speculations as well. He covers not only the well tilled fields, but also such newly cleared research areas as joy and compassion. It is beyond the scope of the present review to evaluate this chapter in detail. While it can doubtless be criticized from many viewpoints, it is effective as a structure for presenting much of the current state of knowledge and thinking in this area, much more so than the chapter in the previous edition.

completely with the author that "child behavior cannot be understood apart from . . . the physical body through which it perceives, reacts, and functions, and to which others react," (p. 292); but unfortunately he finds here no consideration of this problem. Can't we have a chapter on the *psychology* of physical differences in children in the next edition such as we have on the psychology of sex differences in this one?

The chapters of the *Manual* where major changes appear, and the new chapters, may be individually mentioned as follows:

"Methods of Child Psychology" (John Anderson). Anderson wrote chapters with this title in the 1933 edition of Murchison's *Handbook* and the 1946 edition of the *Manual.* These, with the present essay, constitute a revealing exhibit of the rewards of two decades of preoccupation by psychologists with the problems of methodology. Anderson wrote approximately 11,000 words in 1933, 22,000 in 1946, and 33,000 in 1954 and while the increase in level may not be quite so marked as the increase in quantity, the reviewer perceives it as being nearly so. However, the scope of the chapter is too wide; it is, in fact, a survey of general psychological methodology. With a few changes in nomenclature and examples this essay would be as appropriate in a social or an experimental psychology as it is here. One hopes that the time will come when the general field will have been so well covered by others that Anderson can devote his attention to *special* problems of methodology in psychological research with children.

"The Environment and Mental Development" (H. E. Jones). Jones' review shows that there has been an active and profitable period of work on this area. He reports a welcome movement away from global nature-nurture studies to increased specification of the effects of particular environmental influences. Jones appears to take the position, without elaborating and emphasizing it sufficiently, however, that this movement must go still further. He says in connection with the Iowa studies "our chief need in this field is not for statistical methods of greater power and subtility but for more vigorous experimental procedures . . . " (681). If he means here, as the reviewer thinks he does, the need for better definition and control of environmental variables, he has raised a crucial point. In its context, Jones' statement seems to imply that, for example, it is not enough to study the effects of nursery school attendance upon intellectual development; it is necessary to define the significance of nursery school for children in psychologically meaningful concepts: perhaps in such variables as the stimulation, the attractiveness, the freedom, the rewards, and the learning opportunities of the nursery school situation. This points to an important requirement which Jones' review does little to meet. Although his critical evaluations of the technical adequacy of individual studies are excellent, he does little to organize and conceptualize the kinds of problems in this field. All "environmental" influences, ranging from culture, economics, social class, foster homes, hospitalization, and schooling to birth order, season of the year, nutrition, health, physical size, physiological maturity, and race are treated on the same level. Surely the time will soon come when some discrimination and sorting of problems within this field will be possible in terms of primitive theories of environmental variables.

"Character Development in Children—An Objective Approach (Vernon Jones). The author reports many new investigations of relevance to this topic, but the area of understanding appears to have been increased very little in the time since the first edition. It is clear that this is an area where the multiplying of facts is proving of little benefit. New directions are needed.

"Language Development in Children" (McCarthy). This is the longest chapter in the *Manual* (120 pages) with the largest number of references (776). It has been reorganized and greatly expanded for this edition. Extensive new sections cover the vocalization of infants and language disorders in relation to personality development. The problems considered cover a tremendous range including, for example, fetal sounds and the birth cry; vowel and consonant sounds in infant vocalization; vocabulary tests; amount and rate of talking; the function of language in the child's life; the effects of institutionalization and multiple birth upon language development; the correlation between language and motor, intellectual, and social development; language disorder syndromes and personality development; delayed speech; articulatory defects, stuttering. Despite McCarthy's very effective organization and condensation (including two four-page summarizing tables), the amount of material is so overwhelming that one leaves this chapter with a great weariness. Perhaps the basic difficulty is that "spoken language" is in itself not a useful category of behavior; perhaps it is only phenotypic, symptomatic behavior like hand behavior and foot behavior. In any case, to make future reviews more effective, some means of focusing upon the psychologically most meaningful and timely issues is essential. As it stands, this chapter provides a useful guide to a vast literature.

"Psychological Sex Differences" (Terman and Tyler). This chapter has been largely rewritten for this edition of the *Manual*, and it might well serve as a model for the kind of exposition suited to this book: the problem of the chapter is clear, the evidence is not beyond the limits of what can be handled in the space available, the authors make explicit the limits of their review and of the bibliographical sources covered, they include only those publications most relevant to the trend of the evidence as they see it. This method reveals clearly the status of the problems as viewed by the authors. Developments in the interval since the first edition of the *Manual* are more adequate data which confirm earlier findings and which reinforce the authors' earlier caution regarding possible biological and cultural roots of psychological sex differences. Theories and speculations about the possible lines of influence of these factors are not covered. This is a feature of the problem which some might consider of value.

"Emotional Development" (Jersild). This is a completely reorganized and greatly expanded survey of emotionality in children. Jersild places his discussion within an explicitly stated conception of emotion. He considers not only research findings, but various theories and speculations as well. He covers not only the well tilled fields, but also such newly cleared research areas as joy and compassion. It is beyond the scope of the present review to evaluate this chapter in detail. While it can doubtless be criticized from many viewpoints, it is effective as a structure for presenting much of the current state of knowledge and thinking in this area, much more so than the chapter in the previous edition.

"The Adolescent" (Horrocks). It is difficult to conceive of two chapters with this title more different than those of Dennis in 1946 and Horrocks in 1954. Some measure of this difference is found in the fact that while between them they refer to 337 citations of relevant literature dated 1941 or earlier, only 12 per cent of these citations are common to both bibliographies. This must carry an important lesson regarding the preciseness with which *adolescent* is defined and the "reliability" of these two "observers" of the relevant literature. While Dennis defined his problem as that of the behavioral correlates of physiological sexual maturation in humans, Horrocks ranges much more widely. He devotes six pages to the history of studies of adolescence, thirteen pages to the physical and physiological aspects of adolescence, and nine pages to behavior in adolescence. This is a rather elementary, didactic essay on many aspects of adolescent problems.

"Psychopathology of Childhood" (Benda). This is a new chapter in the *Manual;* it covers mental deficiency (30 pages), personality disorders on a higher integral level (eight pages), and childhood schizophrenia (five pages); it has ninety-four citations in the bibliography. It is beyond the reviewer's competence to make evaluative comments on this chapter.

"Social Development" (H. H. and G. L. Anderson). Harold Anderson has, over a number of years, produced an important series of investigations of the social behavior of children. Here for the first time is a vigorous and invigorating statement of the theoretical framework of this research. It is a theory of almost cosmic scope. A sample of the headings used by the Andersons includes: The Characteristics of Biological Growth; Growth of Civilizations; The Place of Value; Social Development and the Second Law of Thermodynamics; Psychological Entropy and Culture; Theory of Probability; "Scientific" Prediction and Organization; Disintegration, Deterioration, and Psychosis; The Growth Circle; The Vicious Circle; Friendliness, Cooperation, and Integrative Behavior; Leadership; Social Learning. The bibliography refers to such diverse scholars as Toynbee, Wiener, Sinnott, Conant, Freud, Spranger, Krech, Dennis, Sears, and Schneirla. The chapter begins with an outline of trends in psychological thinking rooted in Darwin, Preyer, and William James and ends with the recent investigations of Lippitt, Bowlby, and Bronfenbrenner. A theory of such scope is beyond the capacity of this reviewer to evaluate, and the task is not made easier by the fact that forty-five pages are too few to cover all the details and make all the transitions. However, this reader found intriguing the world view so badly sketched. Here is the kind of broad perspective child psychology hopes some day to achieve, and this may be the direction the quest will take. It is to be hoped that this essay can be elaborated elsewhere as it deserves.

The present volume exhibits the same ambiguities as the earlier one regarding its functions. Some chapters are written as for a text; others are scholarly contributions with no concessions to immature students; still others are literature reviews similar to those of the *Psychological Bulletin.* Partly as a consequence of these different conceptions the chapters are uneven in style and level of writing. But this variety is also, without doubt, a product of the uneven development of the different areas. Child psychology is, indeed, in a ragged, unfinished state and the *Manual* reflects this in all its aspects.

One instance of this is found in the way the same material is handled in chapters with overlapping subject matter. When the same issue is considered by more than one author, it is not uncommon to find different interpretations of the data. Another instance is found in the treatment accorded two modern giants of child psychology: Freud and Piaget. Several authors make a valiant effort to include psychoanalytic findings. It is obvious, however, that all are uncomfortable in their efforts to be simultaneously tolerant and scientific. Psychoanalysis in this book is like a foreign body. (One cannot but remark that in this respect the second edition is far ahead of the first, where Freud received little mention, and that the *Manual* is far ahead of most books under a psychoanalytical aegis which give academic psychology even less attention.) Piaget, too, is treated as an alien, for reasons less obvious to the reviewer. There is no reference to Piaget's important recent publications.

This is a good picture of child psychology in 1954. The weaknesses of the *Manual* are largely the weaknesses of the science it surveys. A good manual of child psychology awaits a more mature science of child behavior. In the meantime this book and, it is to be hoped, its future editions provide an important aid in achieving this maturity.

ROGER G. BARKER
University of Kansas

THRALL, R. M., COOMBS, C. H., & DAVIS, R. L. (Eds.) *Decision processes.* New York: John Wiley, 1954. Pp. viii+332. $5.00.

A summer seminar in 1952 brought together economists, mathematicians, psychologists, and a few representatives of other fields, all concerned with some aspect of choice behavior. Each participant communicated the ideas in his field which he regarded as likely to benefit the others, and some of them were stimulated by the interchange to start new research. The 19 heterogeneous papers comprising *Decision Processes* were prepared during the subsequent year. The book makes evident the difficulties in this area and discourages any expectation of early pay-off for the psychologist.

We may contrast two broad lines of interest in decision analysis, the normative and the descriptive. The normative, found in mathematics, statistics, and economics, tries to state how a rational being ought to act. The descriptive is found in psychology, and to an increasing extent in economics. Descriptive studies examine actual choices, and seek a law to predict the choices or a rationalization to "explain" them. Models from the normative studies aid in rationalization, and the descriptions demonstrate, among other things, how far normal behavior is from that of the postulated rational being.

Among the normative papers here, one series compares the various decision criteria (e.g., minimax) which are currently under study. A second series deals with the development, from axioms, of utility functions to describe individual preferences. These papers contain some new mathematical developments. Of most general interest is Marschak's "Towards an Economic Theory of Organization and Information" which deals comprehensively with the value of information in making decisions. This paper is not as useful an introduction, however, as Marschak's chapter in *Mathematical Thinking in the Behavioral Sciences* (Glencoe: Free Press, 1954, P. F. Lazarsfeld, ed.).

On the descriptive side, we have a

variety of experiments on preferences among wagers, on probability learning, and on coalition-forming behavior. The prize paper in this group is a delightfully designed experiment by Hoffman, Lawrence, and Festinger, demonstrating that coalition forming depends not only on the ostensible pay-offs to be obtained, but also on subtle social attitudes among the participants. Bush, Mosteller, and Thompson present a complicated stochastic model for learning in choice situations. Coombs, Raiffa, and Thrall write helpfully on the place of mathematical models in scientific reasoning, and Coombs employs his "ordered metric" in papers on social choice and (with Beardslee) on wagering.

This volume seems more a memorandum among the participants than a presentation for other readers. These reports do not represent substantial, consolidated advances, as all concerned recognize. There is much use of the single-case experiment, or of the tentative mathematical formulation involving assumptions unsatisfactory to the author. The papers are difficult to read, as mathematics and as English. Even Estes, one of the more lucid contributors, writes this appalling sentence: "When the model is interpreted in terms of the present experiment, it turns out that the rate of learning (systematic change in probability of making a given prediction) depends upon the characteristics of the momentary environmental situation but that over a considerable series of trials the probability of making a given prediction tends to a stable asymptotic distribution with the asymptotic mean, for a group of similar individuals run under like conditions, being independent of the momentary environmental situation, in the present experiment, the nature of the signal, S."

Many psychologists are likely to learn something from a few chapters which are closely related to their interests, and to which they can bring considerable background. For the reader wishing a general knowledge of the frontier areas represented here, one would recommend parts of the Lazarsfeld symposium, the recent *Bulletin* paper by Ward Edwards, and Irwin Bross' *Design for Decision* (Macmillan, 1954).

The eventual impact of utility theory on psychology may well be considerable, even though investigators are not yet successful in integrating the field. The concept of preference for various outcome distributions intrudes into the interpretation of learning experiments, where psychologists have generally regarded each reinforcement as having an objective value. Utility theory shows that concepts like *level of aspiration* are oversimplified, and the coalition experiments introduce new and provocative problems for social psychologists. The contrast between actual choice and rational choice has many implications for the psychology of thinking, especially as regards its social determinants. Talented investigators have started to mine in what is evidently a good spot, even if this first load contains little pay dirt.

LEE J. CRONBACH
University of Illinois

MOSES, PAUL J. *The voice of neurosis.* New York: Grune & Stratton, 1954. Pp. v+131. $4.00.

In this volume, the author focuses on attributes of voice, not on verbal communication or speech in the usual sense. Resonance, melody, register illustrate the elements emphasized. Not sharply organized or sufficiently substantiated to serve as a firm introduction to the method or as a convincing summary of research, the book nevertheless opens up intriguing potentialities. Voice patterns reflect peculiarly significant aspects of personality; the neurotic becomes a different voice instrument from the schizophrenic; and therapy combining voice and dynamic elements is indeed a fascinating approach. These are the three themes that run through the book.

The major thesis presents voice as an expressive technique. Just as a graphologist uses handwriting, so Moses uses voice in all its varied aspects to understand personality. He compares a blind analysis with a Rorschach, and integrates other analyses with case material. His ideas about sound approaches to interpretation hold interest for those who use projective or expressive methods. The aspects of voice he emphasizes and his guide lines in analysis suggest that he may have hit on one of the potentially most fruitful areas of expressive behavior—with elements subtly modified throughout important stages of life history and remaining identifiably fixed in adult vocal behavior. With ever increasing fidelity in recording, voice may become a major focus in the understanding of personality.

ROY M. HAMLIN
Western Psychiatric Institute
University of Pittsburgh

NUTTIN, JOSEPH. *Tâche réussite et échec: théorie de la conduite humaine.* Amsterdam: Publications Universitaires de Louvain, 1953. Pp. x+530. 330 fr.

Over a period of about fifteen years, the author of this study has been conducting experiments at the University of Louvain that are intended to show the impact of success and failure upon human personality. They comprise rather simple experiments in perception, learning, and memory. In some experiments, subjects are asked to judge the area of geometrical surfaces, or the number of people shown in pictures. In others, they are asked to recall whether shown two places numbers were added to or subtracted from another such number in a previous showing, or whether words were learned as associated with another word or a number. Arranged in series, the subject is required to render his judgment as each member in the series is presented, and then is told whether his response is right or wrong. In some series the number of successful and unsuccessful judgments are equal, in others one or the other variety predominates. Subsequently, the subject is questioned regarding matters for which the instructions provided no mental set. For the most part, subjects are young people of school and college age. The results of normals are compared with those of manic and melancholic subjects. The experiments are so designed as to be safeguarded against the usual types of psychological error.

Experiments on learning in which some terms are remembered and others forgotten inevitably involve the question as to how successful responses tend to be reinforced. In this context, Nuttin undertakes a critical examination of the meaning of the law of effect. He concludes that Thorndike's view, at least as set forth in his later writings, is too mechanistic. Hull's position is objected to on similar grounds. With respect to the views of Tolman, he is more tolerant. Nuttin expounds a dynamic conception which admits of a cognitive factor in the stimulus-response sequence, and stresses perfection of response rather than repetition of behavior in the learning situation. Reduction of need is designated the essential condition for reinforcement. He finds support for this conclusion in the recent experiments concerning unfinished tasks.

To this reviewer it seems that, having performed numerous experiments on the effect of success and failure upon incidental memory in simple laboratory tasks, Nuttin has undertaken to widen the scope of the bearing of the experiments in such a way as to provide a general discussion of personality dynamics. In an introductory section under the title "Conduct and Result," he gives a critical survey of the chief contributors to experimental and clinical theory in psychology. Here he develops his theory of personality dynamics in which the cumulative effects of success and failure are often so devastating as to leave traumatic effects.

While those who subscribe to strict objectivity in psychology will be apt to deem Nuttin's appeal to a cognitive factor in behavior dynamics as a regression, those who are disposed to stress the bearing of psychology upon everyday life will find much to acclaim. Left as an open question, however, is the degree to which simple laboratory tasks, such as those in which a subject estimates the number of square centimeters in a triangle or the number of people in a picture and is told he is wrong, approximate such real-life failures as loss of academic standing, rejection by an attractive person of the opposite sex, or elimination from the team in sports.

Using a simple style that is free of the more involved forms of idiomatic expression, Nuttin easily conveys not only his ideas but also his enthusiasm for experiment. By integrating both experimental and clinical sources, the work is given a broad practical bearing which will be of interest primarily to specialists in learning and in personality development

MICHAEL J. ZIGLER
Wellesley College

KORNHAUSER, ARTHUR, DUBIN, ROBERT, & ROSS, ARTHUR M. (Eds.) *Industrial conflict.* New York: McGraw-Hill, 1954. Pp. vii+551. $6.00.

The subject of industrial conflict is so broad that one would hope for its treatment to be lengthy, interdisciplinary, and to encompass many specific topics and viewpoints. As such, this book, written under the sponsorship of SPSSI, meets all one's expectations. It sets out with the ambitious purposes of analyzing the determining factors and conditions which give rise to industrial conflict and of assessing various efforts of solution.

The book is divided into five main parts: (*a*) Basic issues concerning industrial conflict. (*b*) Roots of industrial conflict (motivational analysis, organization and leadership of groups in conflict, social and economic influences). (*c*) Dealing with industrial conflict (accommodating to conflict, efforts to remove sources of conflict, social control of industrial conflict). (*d*) Industrial conflict in other societies. (*e*) Industrial conflict; present and future.

Thirty-nine different authors, including academicians, writers, labor leaders, and industrial representatives, have contributed to the book's forty chapters. With such a heterogeneous group, it is inevitable that varying opinions will be found. However, this adds to the book's value by providing a variety of viewpoints and by highlighting some of the unsolved issues. It is exciting to see this interdisciplinary approach being applied increasingly to industrial problems. Certainly such an approach must be satisfying to all "problem-oriented" industrial psychologists who have long recognized that their medicine bag does not always contain the optimum solution, that many of their tasks are not purely psychological in nature but transverse numerous disciplines.

That the book covers a timely topic is obvious—the tables showing man-days idle as a result of strikes are in themselves sufficient proof. That the book asks more questions than it gives answers, that it presents more suggestions for further research than conclusions from past research, is also true. But it also offers a comprehensive picture for the "serious-minded public" who want a general overview of the relations between labor and management groups and among individuals in the industrial setting.

A final word about the authors. In these days when an increasing number of "edited" books turn out to be reprints of journal articles plus a few introductory remarks, it is refreshing to find a book where each author has not only contributed something new to the field but has done so with no monetary remuneration—all proceeds from the book go to the SPSSI.

JEROME H. ELY
Dunlap and Associates, Inc.

SCHAFER, ROY. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954. Pp. xiv+446. $8.75.

Skinner teaches an alert pigeon to peck a bulls-eye in five minutes, by first reinforcing any approximate success. Struggling with a much more challenging puzzle, the book reviewed here is no bulls-eye, but may be hailed with cautious enthusiasm as the most encouraging near miss of its kind yet published.

The author presents a detailed attempt to establish feedback between painstaking observation of complex individual behavior and general "laws." The fact that some of the laws are tentative or dubious need not be emphasized with undue distress. The attempt itself hits at the core of the projective problem. The Rorschach was never a test in the Binet tradition, simplified by design to point up specifically what should be counted. The inkblots call forth behavior which retains a high degree of uniquely individual complexity. To look at such behavior and start counting (*M*, *H*, *D*), or testing fruitless hypotheses, is easy. To tease out a pattern, theme or process that constitutes a meaningful unit is a problem that has baffled both clinicians and statisticians. Those clinicians who seem to have the art have had little success in writing the method. Statisticians like R. B. Cattell disdain the "inventive" response of the projective method, throw up their hands, and say: What we want is a traditional model test of dynamisms!

Schafer's approach involves chiefly three elements: (*a*) a vocabulary or classificatory system taken from the psychoanalytic terminology for ego defenses; (*b*) the use of judgment in teasing out units, with this judgment based on a background of empirical observation, experimental evidence, and thoughtful speculation; and (*c*) a rough check back of these broad units against other empirical evidence (descriptive case material). This still crude approach is not new, but Schafer's formulated approximation represents a step forward in specificity and scope.

The author's bias or biases need not be approved with equal enthusiasm. Specific biases that mar some chapters and some elements throughout the book can be mentioned only with an important reservation: as stated here they represent 90 per cent the reviewer's projections and only 10 per cent the author's attitudinal style. Bluntly, however, on some pages the author does seem to feel that (*a*) the well-adjusted (analyzed) psychologist in a medical setting should accept the healthy masochistic role of a second-class citizen; (*b*) an expert is more someone with wide and approved experience than someone who should be asked to produce expert evidence; (*c*) psychoanalytic theory may not be the ultimate final word, but it is the current final word as far as party line handed down to second-rate citizens is concerned; and (*d*) usually anything the Rorschach reveals is best understood if labeled with a derogatory word (infantile, sadistic, compulsive).

Actually the author struggles consistently against such biases: lauding solid evidence, rejecting Rorschach's fascinating notes on the inkblots as the final word in this area, and recognizing that there is something "fundamentally neurotic" about reporting all observations in terms of derogatory value judgments. Yet his own problems of professional and scientific identity peek through.

He nevertheless succeeds in setting forth the general outline of a process of clinical judgment, or "intuition," that makes sense. Successful judgments may involve the cancelling out of many details based on false assumptions, self-deception, and initially loose speculation. The relaxed acceptance of all elements, good and bad, checked then with critical rigor against general guide lines which Schafer calls theory, may be important. The general feel for such a process is conveyed by Schafer's book. The further analysis of such judgment processes is important to psychology, as a field of study and as a research tool. The pattern of this approach may lead to more fruitful progress than a pattern based from the beginning on an unimaginative reading of the APA's *Technical Recommendations for Psychological Tests.*

ROY M. HAMLIN
Western Psychiatric Institute,
University of Pittsburgh

TAYLOR, W. S. *Dynamic and abnormal psychology.* New York: American Book Co., 1954. Pp. xiv+658. $5.50.

According to a statement in the author's preface, this book was designed primarily as a textbook for courses in abnormal psychology. It is my impression, however, that it is unlikely to win a wide acceptance. My reasons for this judgment follow.

It would appear that this book has grown from Professor Taylor's own course in the subject and no doubt will fit it admirably. However, his course seems rather unique, at least so far as my knowledge of other courses and texts in abnormal psychology goes. It is rare that learning, action, thought, and connector processes are discussed in a course in abnormal psychology and treated there much as they would be in general psychology. Yet five of the 19 chapters deal with these topics, and perhaps half the book is devoted to matters not now ordinarily found in currently widely used texts. As a matter of fact, only one chapter is devoted to the major behavioral disorders—neuroses, psychoses, and certain other categories.

More importantly, the several chapters do not seem to hang together in a compellingly coherent way. Having discussed action, for example, the author seems not to make use of this discussion in later treatments of other topics. Moreover, many of the subjects introduced receive so scanty a discussion as to be unintelligible to the naive reader and simply uninformative to the moderately sophisticated reader. Illustrative, though not representative, is the section entitled, "Psychological Aids to Diagnosis and Prognosis." In two and a fraction pages of actual text, ten topics are discussed, ranging from word association tests and reaction times to projective techniques.

Professor Taylor has obviously read widely, and the book contains numerous quotations from literature and anecdotes of one sort or another drawn from a wide variety of sources. But there is abundance here, to the point often of excessive redundancy, and actual confusion as well as flagging interest is apt to be the result.

Although I feel certain that Professor Taylor has one, the book does not convey to me a theory or a coherent picture of the dynamics of adjustment or of abnormality. It appears to be eclectic, but amorphously so.

It remains to be said that there are many things of value in this volume. The abundant case materials, the many discussions of problems and questions, such, for example, as multiple personality, suggestion, free will, and other matters no longer frequently encountered in our texts, the history of abnormal psychology, and the over-all attempt to employ the categories of general psychology in the special field of dynamics and abnormality are all points of interest and worthy of commendation. As a textbook for general use, however, I do not feel that it can be recommended.

CHARLES N. COFER
University of Maryland

THORPE, LOUIS P., & SCHMULLER, ALLEN M. *Contemporary theories of learning.* New York: Ronald Press Co., 1953. Pp. viii+480. $5.50.

The authors indicate that "It is the purpose of this textbook for university and college students to explain the most important theories of learning in the clearest and simplest possible language, to show the relevance of each of them to the educational process, and to point out that in spite of the many conflicts between these theories they have a common ground upon which can be based an intelligible pattern of classroom procedure" (p. v.) Each aspect of this purpose is in itself a large and important undertaking that makes quite different demands upon the authors and requires different evaluation criteria. Two responsibilities are assumed in accomplishing the first part of their purpose: (*a*) to indicate their criteria for selecting material from a theory, and (*b*) to explain each theory accurately. As the chapters are written the student is likely to infer that the volume presents unabridged theories of seven men (Thorndike, Guthrie, Hull, Skinner, Wheeler, Tolman and Dewey) and two positions—Functionalism and Gestalt. Since no selection criteria were indicated, the authors are open to the criticism that they have promised more than they have delivered. Also, Thorpe and Schmuller have not always been accurate. For example, they failed to differentiate statements about the theorist's program and metatheory from those pertaining to the theory proper. They inferred from Hull's preoccupation with biological survival and adaptation that these are an integral part of his system. Hull, anticipating confusion on this point, wrote that "adaptive considerations are useful in making a preliminary search for postulates, but that once the postulates have been selected they must stand on their own feet." Adaptation intrudes as a persistent theme in their discussions of most theories. For example, they say "By learning Guthrie refers to such behavior (acts) as will assist the organism in making necessary adaptations. From this point of view learning is an additive process in that something helpful to him always accrues to the learner as a result of it" (p. 97). This statement also is incorrect, for Guthrie has indicated that the "commonsense" definition of learning as contrasted with a "scientific" one such as his own "applies learning only to the attainment of good results and we shall find that we acquire bad habits and tendencies to failure in exactly the same way in which we acquire good habits and tendencies to success." The authors in purporting to restate Thorndike's law of effect say, "Stated simply, this law holds that the more one utilizes certain neural pathways—assuming that they exist as realities—the stronger become the bonds" (p. 52).

A fundamental difficulty is inherent in the approach taken by the authors to accomplish their second purpose. At the present time it is not possible to apply entire systems of learning to problems because they are not highly developed, logically integrated sets of axioms and postulates. However, application can take place at the level of the more limited special theories, most of which are components of a general theory.

When special theories are applied, three factors appear to determine their utility: (*a*) the language in which the theory is stated; (*b*) the reference experiments employed by the theorist; and (*c*) the particular use to which the theory is to be put. The first two are the responsibility of the theorist; the third is the choice of the practitioner. Both have neglected their responsibilities. Learning theorists have not always provided adequate interpretations of their theoretical terms, making it impossible to apply parts of their theory in the sense of deducing relationships that should hold among the elements of any new situation. Those who wish to apply an existing special theory generally have not translated the elements and relationships in the new situation into terms used in stating the theory. This is a necessary prerequisite for the accomplishment of the desired rapprochement between theory and practice. Since this was not done by Thorpe and Schmuller, it is not surprising that their final summary of the areas of agreement among learning theorists provides the educator with nothing he did not already know about classroom procedure.

There is another consequence of beginning with well-known "general theories" rather than with a psychological analysis of particular educational problems. Some possible applications were overlooked. Examples include inhibition theory, which could be related to problems of motor skill performance in learning a trade; Gibson's stimulus generalization hypothesis in verbal learning; Cofer and Foley, and also Osgood's mediation hypotheses about language behavior; and Hovland's application of "information theory" to concept learning.

An important oversight is the omission of references to either the tables or figures in the text. Almost none of the 55 figures or 11 tables is referred to in the text.

Lawrence M. Stolurow
University of Illinois

Gruenberg, Sidonie M. (Ed.) *The encyclopedia of child care and guidance.* New York: Doubleday, 1954. Pp. 1016. $7.50.

A most distinguished editorial staff, advisory board, and large roster of contributors were involved in producing this massive book. Part I, the first two-thirds of the book, contains more than 1,000 alphabetically arranged, cross-referenced entries followed by a classified list of agencies and organizations and an annotated list of further readings. Part II consists of 30 excellent chapters on various aspects of child development and the social forces affecting children.

The encyclopedia portion of the book is consistently addressed to the stereotype of a literate but completely naive parent. The style is chatty and nontechnical while the pages are decorated with an abundance of line drawings, mostly of children. Inasmuch as almost any topic is likely to appear (e.g., Pediatrician, Pediatrics Clinic, Pediculosis Capitis, Peek-A-Boo, Pellagra, Pelvic Examination, Pelvis, Penicillin, Percussion Instruments, Period, Periodicals for Children, Permanent Waves, etc.), the book is guaranteed to furnish even the most professional reader with interesting bits of information. Generally speaking, however, Part II will be of greater interest to psychologists as a popular but comprehensive review of current thinking.

Certainly few will quarrel with the repetitive message: Be loving and patient but seek professional help when a real problem exists. The reviewer foresees little competition for Dr. Spock (himself a contributor) at 35¢ or even Dr. Carmichael at $12.00.

LEONARD S. KOGAN
Institute of Welfare Research
Community Service Society of New York

SONNEMANN, ULRICH. *Existence and therapy.* New York: Grune & Stratton, 1954. Pp. xi+372. $5.00.

This will doubtlessly prove to be an exceptionally difficult book for the average psychologist to understand, whether he be clinician or experimentalist. Not only is Dr. Sonnemann's manner of writing exquisitely involved; but his ideas, aside from the violent emotional reactions they are bound to arouse in most scientifically minded readers, almost defy any purely intellectual comprehension. In Dr. Sonnemann's own existentialist view, these ideas, like all human knowledge, must presumably be experienced-as-being, or known-in-themselves, before they can be factually analyzed or "understood." They must be accepted, as it were, on faith—faith in man's being or existence. This may be so; and certainly logical positivists, who are positively anathema to dyed-in-the-wool existentialists, would never argue that it may not be so, but merely that the question of whether or not it may be so, being essentially unprovable by empirical observation, is a meaningless one. The question is: Why, believing as he does that truth or knowledge can only be directly experienced by man as part of his essential humanity or being, does Dr Sonnemann (and his fellow existentialists) go to the trouble of long-windedly explaining to his fellow psychologists, in what suspiciously appears to be a highly analytic manner, why existentialism is far superior to every other system of psychology and philosophy ever known to man?

Existence and Therapy is a thoroughgoing discussion of the existentialist viewpoint of Heidegger, Jaspers, Binswanger, Boss, and other recent European philosophers and clinicians. It attempts to make mincemeat of empiricism, objectivism, experimentalism, Freudianism, Gestaltism, Jungianism, and virtually every other influential contemporary way of thinking and therapizing. In so doing, it makes some telling points, particularly in relation to some of the shortcomings of orthodox Freudians and Gestaltists.

It is on the more positive side that Dr. Sonnemann's detailed presentation of daseinsanalytic theories seems to have much less to offer. Where these theories overlap with neopsychoanalytic and holistic views of man as a total autonomous-autochthonous organism who ceaselessly interacts with other human beings and the external world, and who must be understood and therapized in the light of his biological-social origins and development, existentialism seems eminently sane. But where it goes off, and often, into concepts of being and the naught, existence as being-toward-death, the who of existence, the universality of the love norm, etc., daseinsanalysis appears to become, at least to this empiricism-biased reviewer, distinctly unobjective, mystical, moralistic, and tautological. Its criticisms of other viewpoints sometimes make good sense; but its own espousals often seem to be double-talk.

In any event, Dr. Sonnemann has herewith given us the first quite detailed and reasonably definitive English-language explanation of the daseinsanalytic philosophy. For all its difficulty and blood-pressure raising potentiality, his book is replete with provocative thoughts that certainly deserve a hearing.

Albert Ellis

New York City

Pennington, L. A., & Berg, Irwin A. (Eds.) *An introduction to clinical psychology.* (2nd Ed.) New York: Ronald Press, 1954. Pp. viii+709. $6.50.

The student introduced to clinical psychology through this excellent revision of an established text would be well introduced indeed and he should be both impressed and attracted to further study. He certainly could conclude that clinical psychologists are concerned with many things, from myopia to ethics. He should gain assurance of finding a compatible role in the field whether he is attracted most by the opportunity to carry out research or to apply an art in the interest of humanity. He would, of course, be forewarned that the clinical psychologist is *expected* to burn both ends of the science-service candle, but it will be apparent to him from reading the text that his mentors often burn more brightly at one end than the other.

The involved student might have trouble with Cattell's introductory chapter, particularly if he rereads it after seeing what the other 29 authors have to say. Cattell doesn't seem to know his own mind, and his ideas sometimes seem captive to his phrases. He states that the clinical psychologist should limit his concern to "those who for various reasons have failed in life's educational process" (p. 4), and that "the essential purpose of the profession is to guide the sick back to health" (p. 21). Thus the "ideal practitioner" of clinical psychology should have medical training in addition to his doctoral study in psychology. Finally, for "industrial personnel work, vocational guidance, and educational psychology, the basic and central type of training and qualification needed is that of the clinical psychologist" (p. 20). He is impressed by the derivation of "clinical" from "bed," and it is left to the other authors to disregard etymology and provide operational definitions indicating a much wider scope for clinical psychology. Furthermore the authors of chapters 2 to 25 seem unaware that they are to be saved by factor analysis. A suggestion to the editors in the third edition: keep Cattell, for he is always stimulating, but get two or three others to write introductory chapters entitled "The Meaning of Clinical Psychology."

This second edition is substantially a new book. A hundred pages have been added. Ten new chapters have been included, six have been dropped, and the remaining revised considerably. Notable improvement has been achieved in presenting treatment procedures. The editors have sought unity in a volume likely to fall apart by encouraging the authors to keep problem centered rather than technique centered. Most authors succeed in doing this; the chapters by Shoben, Mowrer, Garner, Dorcus, and Pennington come to mind. Berg gets involved in technique to the point of warning against the distraction in an interview of a cluttered desk or of bright objects in the border of the visual field. Sargent and Hirsch supply a solid chapter on projective methods, possibly more useful to doctoral students reviewing for exams than intriguing to the neophyte. In this edition, a chapter on the professional relationships of the clinical psychologist is replaced by one on research. If the choice is between the two, the focus on research is certainly to be preferred. But it would be useful to the beginning student to know something of the professional problems and responsibilities of the clinical psychologist.

The writing is good throughout the book. Some chapters seem a bit ponderous (like that of Saslow, Guze, and Matarazzo), some simple and highly effective (like that of McCandless), and some brilliant (like that of Mowrer). A great deal of care has gone into the selection of the bibliographies. They should be genuinely useful to the beginning student. The publishers might have been more generous in their provisions for illustrations, since introductions are better remembered, or at least enjoyed more, when the subject is charming to the eye. Clinical psychology has more to offer visually than visual acuity test charts. Biographical notes on the authors, highlighting their work and thus exhibiting what clinical psychologists do, might add to the value and interest of the book to the beginning student.

Because of the diligence of the editors and the excellence of the authors, clinical psychology now has an even better introductory text than the good one it has had for the past six years.

Nicholas Hobbs

Peabody College

King, H. E. *Psychomotor aspects of mental disease: an experimental study*. Cambridge: Harvard Univer. Press, 1954. Pp. xiv+185. $3.50.

The current emphasis on projective techniques has unfortunately tended to restrict experimental investigations utilizing other procedures in the study of mental patients. In demonstrating a close relation between severity of behavior disorder and psychomotor performance, as measured by such "old-fashioned" tests as reaction time, speed of tapping, and finger dexterity, King has contributed a quantified technique for measuring the current mental status of the patient. Indirectly, his findings emphasize the importance of the "open" approach in psychodiagnostic work. In the long run progress depends on the development of new techniques rather than on the ultrarefinement of available tools.

James D. Page

Temple University

THELEN, HERBERT A. *Dynamics of groups at work.* Chicago: Univer. of Chicago Press, 1954. Pp. x+379. $6.00

The study of the group has always been a major and is still an increasing concern of social psychology and sociology. And a great many social psychologists (some labeled psychologists and some sociologists) have contributed to what we now know concerning the nature of the human group. Among more recent contributors, Moreno, Lewin, and their followers have been prominent, even though their contributions are not quite as unique and fundamental as their followers persist in claiming.

The intellectual offspring of Lewin include not only social psychologists, who have added to our knowledge of groups and who have shown useful ways to apply this information to the solution of practical problems, but also a growing social engineering cult. This cult has taken and sanctified the words of Lewin and has attributed to him and his followers not only their own substantial findings but also other theories and techniques common among organizational and human relations specialists. These cult members brush up, simplify, and give freshened terminology to such items in order to make them readily saleable to civic leaders, social work directors, education administrators, and business managers. They offer a new dispensation from those mysterious creatures, the "scientific social psychologists." Thelen's book is a new and rather comprehensive exegesis on this cult's materials.

A science finds its application in the form of technologies. Persons functioning as technicians or engineers carry scientific findings into shops and neighborhoods and clinics and board rooms. And as the recent history of specialists in the various fields demonstrates, engineers quickly seek professional reinforcement in cult-like groups and organizations. Such groups have a pressing "mission" not too well understood and appreciated by the uninitiated or even by the scientists from whom their knowledge derives. They also have initiation requirements and rites, revered leaders, and technical skills dependably applied only by cult members. For the cult to grow sturdily, too, through the increasing production of recognized and recognizable members, its materials must be standardized and rather precisely communicable. As Thelen demonstrates by his book, the group dynamics cult has all these characteristics.

The unique efforts of the scientist flourish best in the hands of individualists, but the fate of technicians and engineers is a joint one. It depends to a large degree upon joint success in jockeying their secular order of specialists into positions of prestige, control, and thus power. In areas where specialties more obviously overlap and thus more sharply compete, such as in human relations, such joint cult-like efforts to cope with competition become all the more mandatory.

This development of a "new" kit of "tools" for management apparently results not only in the caricaturing of tested social psychological theories but also in a striking switch in basic values served. The caricaturing is evinced in Thelen's twelve lists of guidance principles for group workers and groups. These are based far more upon an unquestioning acceptance of a typical middle class societal surrogate's version of current societal morality than upon scientific findings concerning groups. The bare mention of social classes and the failure to grasp their significance in personality and in group functioning is a part of this problem. The switch in values appears in recurrent emphases by Thelen upon manipulation to achieve morally preconceived goals (rather than upon scientific discovery of the nature of group members, goals, and processes) and upon the service of those concerned with societal stability (rather than upon the stimulation of democratic processes of societal change).

Thelen's first six chapters outline and illustrate six "technologies," which, in his view, while differing, have behind them "fundamental similarities." All six are apparently endorsed. Yet in their outlines, one finds such statements as these: Behavior "is relevant and useful when it contributes to getting the job done, and it is hindering or nonuseful to the extent that it is a response to ideological, class, or racial factors." (Thelen apparently stereotypes "ideological, class, or racial factors"; he does not reveal an appreciation of their implications.) "Neighbors were no longer seen as people one could identify with; the sense of common cause was lost." (He is handicapped by sentimental connotations of the word, neighborhood.) "The role of the student is determined by the will of the teacher—with some qualification by the standards of the peer and family groups to which the student belongs." (Thelen apparently has not studied how peer and family group participation orients participation in subsequent groups and how class and ethnoid factors alter such participation patterns.)

As I studied this book, I constantly found myself seeing an authoritarian dean of students read it, nod his head in satisfaction, and then go forth and quote formulas from it to justify the disruption of democratic student experiments. When the princes had such as Machiavelli, propaganda analysts for the common man had it difficult enough. Now that they have such polished social engineers as Thelen who sincerely work "for the development of the 'humane community' toward which man's nature . . . is driving him," analysis becomes all the more difficult and efforts to communicate it all the more confusing. But I trust I am overrating the group dynamics cult! Fortunately society has many antitoxins against its continued successful manipulation. Some of them may work slowly, but they work.

ALFRED McCLUNG LEE

Brooklyn College

EDITORIAL NOTE

As approved by the Board of Directors and the Council of the APA, beginning with January 1956, a new journal will be published by the Association. This journal will review books, monographs, films and related publications—a function presently performed by four different APA journals.

Accordingly, book reviews will appear in the *Psychological Bulletin* only through the completion of the present volume, **52**, November 1955 issue.

Hereafter all publications submitted for review and requests to prepare reviews should be directed to the editor of the new journal, *Contemporary Psychology, A Journal of Reviews:*

E. G. Boring, Editor
Memorial Hall
Harvard University
Cambridge 38, Mass.

The editors wish to take this opportunity to express their considerable gratitude to the cooperating book reviewers. Any contribution that this department may have made can be attributed only to the work of the reviewers, which we hope has not been a case of love's labour lost.

E. G.
W. D.

VOL. 52, NO. 4 JULY, 1955

Psychological Bulletin

CONSTRUCT VALIDITY IN PSYCHOLOGICAL TESTS

LEE J. CRONBACH
University of Illinois

AND

PAUL E. MEEHL[1]
University of Minnesota

Validation of psychological tests has not yet been adequately conceptualized, as the APA Committee on Psychological Tests learned when it undertook (1950–54) to specify what qualities should be investigated before a test is published. In order to make coherent recommendations the Committee found it necessary to distinguish four types of validity, established by different types of research and requiring different interpretation. The chief innovation in the Committee's report was the term *construct validity*.[2] This idea was first formulated by a subcommittee (Meehl and R. C. Challman) studying how proposed recommendations would apply to projective techniques, and later modified and clarified by the entire Committee (Bordin, Challman, Conrad, Humphreys, Super, and the present writers). The statements agreed upon by the Committee (and by committees of two other associations) were published in the *Technical Recommendations* (**59**). The present interpretation of construct validity is not "official" and deals with some areas where the Committee would probably not be unanimous. The present writers are solely responsible for this attempt to explain the concept and elaborate its implications.

Identification of construct validity was not an isolated development. Writers on validity during the preceding decade had shown a great deal of dissatisfaction with conventional notions of validity, and introduced new terms and ideas, but the resulting aggregation of types of validity seems only to have stirred the muddy waters. Portions of the distinctions we shall discuss are implicit in Jenkins' paper, "Validity for what?" (**33**), Gulliksen's "Intrinsic validity" (**27**), Goodenough's distinction between tests as "signs" and "samples" (**22**), Cronbach's separation of "logical" and "empirical" validity (**11**), Guilford's "factorial validity" (**25**), and Mosier's papers on "face validity" and "validity generalization" (**49, 50**). Helen Peak (**52**) comes close to an explicit statement of construct validity as we shall present it.

FOUR TYPES OF VALIDATION

The categories into which the *Recommendations* divide validity studies are: predictive validity, concurrent validity, content validity, and construct validity. The first two of these may be considered together as *criterion-oriented* validation procedures.

The pattern of a criterion-oriented

---

[1] The second author worked on this problem in connection with his appointment to the Minnesota Center for Philosophy of Science. We are indebted to the other members of the Center (Herbert Feigl, Michael Scriven, Wilfrid Sellars), and to D. L. Thistlethwaite of the University of Illinois, for their major contributions to our thinking and their suggestions for improving this paper.

[2] Referred to in a preliminary report (**58**) as *congruent validity*.

study is familiar. The investigator is primarily interested in some criterion which he wishes to predict. He administers the test, obtains an independent criterion measure on the same subjects, and computes a correlation. If the criterion is obtained some time after the test is given, he is studying *predictive validity*. If the test score and criterion score are determined at essentially the same time, he is studying *concurrent validity*. Concurrent validity is studied when one test is proposed as a substitute for another (for example, when a multiple-choice form of spelling test is substituted for taking dictation), or a test is shown to correlate with some contemporary criterion (e.g., psychiatric diagnosis).

Content validity is established by showing that the test items are a sample of a universe in which the investigator is interested. Content validity is ordinarily to be established deductively, by defining a universe of items and sampling systematically within this universe to establish the test.

Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not "operationally defined." The problem faced by the investigator is, "What constructs account for variance in test performance?" Construct validity calls for no new scientific approach. Much current research on tests of personality (**9**) is construct validation, usually without the benefit of a clear formulation of this process.

Construct validity is not to be identified solely by particular investigative procedures, but by the orientation of the investigator. Criterion-oriented validity, as Bechtoldt emphasizes (**3**, p. 1245), "involves the *acceptance* of a set of operations as an adequate definition of whatever is to be measured." When an investigator believes that no criterion available to him is fully valid, he perforce becomes interested in construct validity because this is the only way to avoid the "infinite frustration" of relating every criterion to some more ultimate standard (**21**). In content validation, *acceptance* of the universe of content as defining the variable to be measured is essential. Construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured. Determining what psychological constructs account for test performance is desirable for almost any test. Thus, although the MMPI was originally established on the basis of empirical discrimination between patient groups and so-called normals (concurrent validity), continuing research has tried to provide a basis for describing the personality associated with each score pattern. Such interpretations permit the clinician to predict performance with respect to criteria which have not yet been employed in empirical validation studies (cf. **46**, pp. 49–50, 110–111).

We can distinguish among the four types of validity by noting that each involves a different emphasis on the criterion. In predictive or concurrent validity, the criterion behavior is of concern to the tester, and he may have no concern whatsoever with the type of behavior exhibited in the test. (An employer does not care if a worker can manipulate blocks, but the score on the block test may predict something he cares about.) Content validity is studied when the tester *is* concerned with the type of behavior involved in the test performance. Indeed, if the test is a work sample, the behavior represented in the test may be an end in itself. Construct validity is ordinarily studied when the tester has no definite criterion measure of the quality with which he is concerned, and must use indirect measures. Here the trait or quality underlying the test is of central importance, rather than either the test behavior or the scores on the criteria (**59**, p. 14).

Construct validation is important at times for every sort of psychological test: aptitude, achievement, interests, and so on. Thurstone's statement is interesting in this connection:

> In the field of intelligence tests, it used to be common to define validity as the correlation between a test score and some outside criterion. We have reached a stage of sophistication where the test-criterion correlation is too coarse. It is obsolete. If we attempted to ascertain the validity of a test for the second space-factor, for example, we would have to get judges [to] make reliable judgments about people as to this factor. Ordinarily their [the available judges'] ratings would be of no value as a criterion. Consequently, validity studies in the cognitive functions now depend on criteria of internal consistency . . . (60, p. 3).

Construct validity would be involved in answering such questions as: To what extent is this test of intelligence culture-free? Does this test of "interpretation of data" measure reading ability, quantitative reasoning, or response sets? How does a person with A in Strong Accountant, and B in Strong CPA, differ from a person who has these scores reversed?

Example of construct validation procedure. Suppose measure X correlates .50 with Y, the amount of palmar sweating induced when we tell a student that he has failed a Psychology I exam. Predictive validity of X for Y is adequately described by the coefficient, and a statement of the experimental and sampling conditions. If someone were to ask, "Isn't there perhaps another way to interpret this correlation?" or "What other kinds of evidence can you bring to support your interpretation?", we would hardly understand what he was asking because no interpretation has been made. These questions become relevant when the correlation is advanced as evidence that "test X measures anxiety proneness." Alternative interpretations are possible; e.g., perhaps the test measures "academic aspiration," in which case we will expect different results if we induce palmar sweating by economic threat. It is then reasonable to inquire about other *kinds* of evidence.

Add these facts from further studies: Test X correlates .45 with fraternity brothers' ratings on "tenseness." Test X correlates .55 with amount of intellectual inefficiency induced by painful electric shock, and .68 with the Taylor Anxiety scale. Mean X score decreases among four diagnosed groups in this order: anxiety state, reactive depression, "normal," and psychopathic personality. And palmar sweat under threat of failure in Psychology I correlates .60 with threat of failure in mathematics. Negative results eliminate competing explanations of the X score; thus, findings of negligible correlations between X and social class, vocational aim, and value-orientation make it fairly safe to reject the suggestion that X measures "academic aspiration." We can have substantial confidence that X does measure anxiety proneness if the current theory of anxiety can embrace the variates which yield positive correlations, and does not predict correlations where we found none.

KINDS OF CONSTRUCTS

At this point we should indicate summarily what we mean by a construct, recognizing that much of the remainder of the paper deals with this question. A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct. We expect a person at any time to possess or not possess a qualitative attribute (amnesia) or structure, or to possess some degree of a quantitative attribute (cheerfulness). A construct has certain associated meanings carried in statements of this general character: Persons who possess this attribute will, in situation X, act in manner Y (with a stated probability). The logic of construct validation is invoked whether the construct is highly systematized or loose, used in ramified theory or a few simple propositions, used in absolute propositions or probability statements. We seek to specify how one is to defend a proposed interpretation of a test; *we are not recommending any one type of interpretation.*

The constructs in which tests are to be interpreted are certainly not likely to be physiological. Most often they will be traits such as "latent hostility" or "variable in mood," or descriptions in terms of an educational objective, as "ability to plan experiments." For the benefit of readers who may have been influenced by certain eisegeses of MacCorquodale and Meehl (40), let us here emphasize: Whether or not an interpretation of a test's properties or relations involves questions of construct validity is to be decided by examining the entire body of evidence offered, together with what is asserted about the test in the context of this evidence. Proposed identifications of constructs allegedly measured by the test with constructs of other sciences (e.g., genetics, neuroanatomy, biochemistry) make up only *one* class of construct-validity claims, and a rather minor one at present. Space does not permit full analysis of the relation of the present paper to the MacCorquodale-Meehl distinction between hypothetical constructs and intervening variables. The philosophy of science pertinent to the present paper is set forth later in the section entitled, "The nomological network."

THE RELATION OF CONSTRUCTS TO "CRITERIA"

Critical View of the Criterion Implied

An unquestionable criterion may be found in a practical operation, or may be established as a consequence of an operational definition. Typically, however, the psychologist is unwilling to use the directly operational approach because he is interested in building theory about a generalized construct. A theorist trying to relate behavior to "hunger" almost certainly invests that term with meanings other than the operation "elapsed-time-since-feeding." If he is concerned with hunger as a tissue need, he will not accept time lapse as *equivalent* to his construct because it fails to consider, among other things, energy expenditure of the animal.

In some situations the criterion is no more valid than the test. Suppose, for example, that we want to know if counting the dots on Bender-Gestalt figure five indicates "compulsive rigidity," and take psychiatric ratings on this trait as a criterion. Even a conventional report on the resulting correlation will say something about the extent and intensity of the psychiatrist's contacts and should describe his qualifications (e.g., diplomate status? analyzed?).

Why report these facts? Because data are needed to indicate whether the criterion is any good. "Compulsive rigidity" is not really intended to mean "social stimulus value to psychiatrists." The implied trait involves a range of behavior-dispositions which may be very imperfectly sampled by the psychiatrist. Suppose dot-counting does not occur in a particular patient and yet we find that the psychiatrist has rated him as "rigid." When questioned the psychiatrist tells us that the patient was a rather easy, free-wheeling sort;

however, the patient *did* lean over to straighten out a skewed desk blotter, and this, viewed against certain other facts, tipped the scale in favor of a "rigid" rating. On the face of it, counting Bender dots may be just as good (or poor) a sample of the compulsive-rigidity domain as straightening desk blotters is.

Suppose, to extend our example, we have four tests on the "predictor" side, over against the psychiatrist's "criterion," and find generally positive correlations among the five variables. Surely it is artificial and arbitrary to impose the "test-should-predict-criterion" pattern on such data. The psychiatrist samples verbal content, expressive pattern, voice, posture, etc. The psychologist samples verbal content, perception, expressive pattern, etc. Our proper conclusion is that, from this evidence, the four tests and the psychiatrist all assess some common factor.

The asymmetry between the "test" and the so-designated "criterion" arises only because the terminology of predictive validity has become a commonplace in test analysis. In this study where a construct is the central concern, any distinction between the merit of the test and criterion variables would be justified only if it had already been shown that the psychiatrist's theory and operations were excellent measures of the attribute.

INADEQUACY OF VALIDATION IN TERMS OF SPECIFIC CRITERIA

The proposal to validate constructual interpretations of tests runs counter to suggestions of some others. Spiker and McCandless (57) favor an operational approach. Validation is replaced by compiling statements as to how strongly the test predicts other observed variables of interest. To avoid requiring that each new variable be investigated completely by itself, they allow two variables to collapse into one whenever the properties of the operationally defined measures are the same: "If a new test is demonstrated to predict the scores on an older, well-established test, then an evaluation of the predictive power of the older test may be used for the new one." But accurate inferences are possible only if the two tests correlate so highly that there is negligible reliable variance in either test, independent of the other. Where the correspondence is less close, one must either retain all the separate variables operationally defined or embark on construct validation.

The practical user of tests must rely on constructs of some generality to make predictions about new situations. Test X could be used to predict palmar sweating in the face of failure without invoking any construct, but a counselor is more likely to be asked to forecast behavior in diverse or even unique situations for which the correlation of test X is unknown. Significant predictions rely on knowledge accumulated around the generalized construct of anxiety. The *Technical Recommendations* state:

> It is ordinarily necessary to evaluate construct validity by integrating evidence from many different sources. The problem of construct validation becomes especially acute in the clinical field since for many of the constructs dealt with it is not a question of finding an imperfect criterion but of finding any criterion at all. The psychologist interested in construct validity for clinical devices is concerned with making an estimate of a hypothetical internal process, factor, system, structure, or state and cannot expect to find a clear unitary behavioral criterion. An attempt to identify any one criterion measure or any composite as *the* criterion aimed at is, however, usually unwarranted (59, p. 14–15).

This appears to conflict with arguments for specific criteria prominent at places in the testing literature.

Thus Anastasi (**2**) makes many statements of the latter character: "It is only as a measure of a specifically defined criterion that a test can be objectively validated at all . . . To claim that a test measures anything over and above its criterion is pure speculation" (p. 67). Yet elsewhere this article supports construct validation. Tests can be profitably interpreted if we "know the relationships between the tested behavior . . . and other behavior samples, none of these behavior samples necessarily occupying the preeminent position of a criterion" (p. 75). Factor analysis with several partial criteria might be used to study whether a test measures a postulated "general learning ability." If the data demonstrate specificity of ability instead, such specificity is "useful in its own right in advancing our knowledge of behavior; it should not be construed as a weakness of the tests" (p. 75).

We depart from Anastasi at two points. She writes, "The validity of a psychological test should not be confused with an analysis of the factors which determine the behavior under consideration." We, however, regard such analysis as a most important type of validation. Second, she refers to "the will-o'-the-wisp of psychological processes which are distinct from performance" (**2**, p. 77). While we agree that psychological processes are elusive, we are sympathetic to attempts to formulate and clarify constructs which are evidenced by performance but distinct from it. Surely an inductive inference based on a pattern of correlations cannot be dismissed as "pure speculation."

Specific Criteria Used Temporarily: The "Bootstraps" Effect

Even when a test is constructed on the basis of a specific criterion, it may ultimately be judged to have greater construct validity than the criterion. We start with a vague concept which we associate with certain observations. We then discover empirically that these observations covary with some other observation which possesses greater reliability or is more intimately correlated with relevant experimental changes than is the original measure, or both. For example, the notion of temperature arises because some objects feel hotter to the touch than others. The expansion of a mercury column does not have face validity as an index of hotness. But it turns out that (*a*) there is a statistical relation between expansion and sensed temperature; (*b*) observers employ the mercury method with good interobserver agreement; (*c*) the regularity of observed relations is increased by using the thermometer (e.g., melting points of samples of the same material vary little on the thermometer; we obtain nearly linear relations between mercury measures and pressure of a gas). Finally, (*d*) a theoretical structure involving unobservable microevents—the kinetic theory—is worked out which explains the relation of mercury expansion to heat. This whole process of conceptual enrichment begins with what in retrospect we see as an extremely fallible "criterion"—the human temperature sense. That original criterion has now been relegated to a peripheral position. We have lifted ourselves by our bootstraps, but in a legitimate and fruitful way.

Similarly, the Binet scale was first valued because children's scores tended to agree with judgments by schoolteachers. If it had not shown this agreement, it would have been discarded along with reaction time and the other measures of ability previously tried. Teacher judgments once constituted the criterion against which the individual intelligence test was validated. But if today a child's IQ is 135 and three of his teachers complain about how stupid he is, we do not conclude that the test has failed. Quite to the contrary, if no error in test procedure can be argued, we treat the test score as a valid statement about an important quality, and define our task as that of finding out what other variables—personality, study skills, etc.—modify achievement or distort teacher judgment.

EXPERIMENTATION TO INVESTIGATE CONSTRUCT VALIDITY

Validation Procedures

We can use many methods in construct validation. Attention should particularly be drawn to Macfarlane's survey of these methods as they apply to projective devices (41).

Group differences. If our understanding of a construct leads us to expect two groups to differ on the test, this expectation may be tested directly. Thus Thurstone and Chave validated the Scale for Measuring Attitude Toward the Church by showing score differences between church members and nonchurchgoers. Churchgoing is not *the* criterion of attitude, for the purpose of the test is to measure something other than the crude sociological fact of church attendance; on the other hand, failure to find a difference would have seriously challenged the test.

Only coarse correspondence between test and group designation is expected. Too great a correspondence between the two would indicate that the test is to some degree invalid, because members of the groups are expected to overlap on the test. Intelligence test items are selected initially on the basis of a correspondence to age, but an item that correlates .95 with age in an elementary school sample would surely be suspect.

Correlation matrices and factor analysis. If two tests are presumed to measure the same construct, a correlation between them is predicted. (An exception is noted where some second attribute has positive loading in the first test and negative loading in the second test; then a low correlation is expected. This is a testable interpretation provided an external measure of either the first or the second variable exists.) If the obtained correlation departs from the expectation, however, there is no way to know whether the fault lies in test A, test B, or the formulation of the construct. A matrix of intercorrelations often points out profitable ways of dividing the construct into more meaningful parts, factor analysis being a useful computational method in such studies.

Guilford (26) has discussed the place of factor analysis in construct validation. His statements may be extracted as follows:

"The personnel psychologist wishes to know 'why his tests are valid.' He can place tests and practical criteria in a matrix and factor it to identify 'real dimensions of human personality.' A factorial description is exact and stable; it is economical in explanation; it leads to the creation of pure tests which can be combined to predict complex behaviors." It is clear that factors here function as constructs. Eysenck, in his "criterion analysis" (18), goes farther than Guilford, and shows that factoring can be used explicitly to test hypotheses about constructs.

Factors may or may not be weighted with surplus meaning. Certainly when they are regarded as "real dimensions" a great deal of surplus meaning is implied, and the interpreter must shoulder a substantial burden of proof. The alternative view is to regard factors as defining a working reference frame, located in a convenient manner in the "space" defined by all behaviors of a given type. Which set of factors from a given matrix is "most useful" will depend partly on predilections, but in essence the best construct is the one around which we can build the greatest number of inferences, in the most direct fashion.

Studies of internal structure. For many constructs, evidence of homogeneity within the test is relevant in judging validity. If a trait such as *dominance* is hypothesized, and the items inquire about behaviors subsumed under this label, then the hypothesis appears to require that these items be generally intercorrelated. Even low correlations, if consistent, would support the argument that people may be fruitfully described in terms of a generalized tendency to dominate or not dominate. The general quality would have power to predict behavior in a variety of situations represented by the specific items. Item-test correlations and certain reliability formulas describe internal consistency.

It is unwise to list uninterpreted data of this sort under the heading "validity" in test manuals, as some authors have done. High internal consistency may *lower* validity. Only if the underlying theory of the trait being measured calls for high item intercorrelations do the correlations support construct validity. Negative item-test correlations may support construct validity, provided that the items with negative correlations are believed irrelevant to the postulated construct and serve as suppressor variables (**31**, p. 431–436; **44**).

Study of distinctive subgroups of items within a test may set an upper limit to construct validity by showing that irrelevant elements influence scores. Thus a study of the PMA space tests shows that variance can be partially accounted for by a response set, tendency to mark many figures as similar (**12**). An internal factor analysis of the PEA Interpretation of Data Test shows that in addition to measuring reasoning skills, the test score is strongly influenced by a tendency to say "probably true" rather than "certainly true," regardless of item content (**17**). On the other hand, a study of item groupings in the DAT Mechanical Comprehension Test permitted rejection of the hypothesis that knowledge about specific topics such as gears made a substantial contribution to scores (**13**).

Studies of change over occasions. The stability of test scores ("retest reliability," Cattell's "N-technique") may be relevant to construct validation. Whether a high degree of stability is encouraging or discouraging for the proposed interpretation depends upon the theory defining the construct.

More powerful than the retest after uncontrolled intervening experiences is the retest with experimental intervention. If a transient influence swings test scores over a wide range, there are definite limits on the extent to which a test result can be interpreted as reflecting the typical behavior of the individual. These are examples of experiments which have indicated upper limits to test validity: studies of differences associated with the examiner in projective testing, of change of score under alternative directions ("tell the truth" vs. "make yourself look good to an employer"), and of coachability of mental tests. We may recall Gulliksen's distinction (**27**): When the coaching is of a sort that improves the pupil's intellectual functioning in school, the test which is affected by the coaching has validity as a measure of intellectual functioning; if the coaching improves test taking but not school performance, the test which responds to the coaching has poor validity as a measure of this construct.

Sometimes, where differences between individuals are difficult to assess by any means other than the test, the experimenter validates by determining whether the test can detect induced intra-individual differences. One might hypothesize that the Zeigarnik effect is a measure of ego involvement, i.e., that with ego involvement there is more recall of incomplete tasks. To support such an interpretation, the investigator will try to induce ego involvement on some task by appropriate directions and compare subjects' recall with their recall for tasks where there was a contrary induction. Sometimes the intervention is drastic. Porteus finds (**53**) that brain-operated patients show disruption of performance on his maze, but do not show impaired performance on conventional verbal tests and argues therefrom that his test is a better measure of planfulness.

Studies of process. One of the best ways of determining informally what accounts for variability on a test is the observation of the person's process of performance. If it is supposed, for example, that a test measures mathematical competence, and yet observation of students' errors shows that erroneous reading of the question is common, the implications of a low score are altered. Lucas in this way showed that the Navy Relative Movement Test, an aptitude test, actually involved two different abilities: spatial visualization and mathematical reasoning (**39**).

Mathematical analysis of scoring procedures may provide important negative evidence on construct validity. A recent analysis of "empathy" tests is perhaps worth citing (**14**). "Empathy" has been operationally defined in many studies by the ability of a judge to predict what responses will be given on some questionnaire by a subject he has observed briefly. A mathematical argument has shown, however, that the scores depend on several attributes of the judge which enter into his perception of *any* individual, and that they therefore cannot be interpreted as evidence of his ability to interpret cues offered by particular others, or his intuition.

The Numerical Estimate of Construct Validity

There is an understandable tendency to seek a "construct validity coefficient." A numerical statement of the degree of construct validity would be a statement of the proportion of the test score variance that is attributable to the construct variable. This numerical estimate can sometimes be arrived at by a factor analysis, but since present methods of factor analysis are based on linear relations, more general methods will ultimately be needed to deal with many quantitative problems of construct validation.

Rarely will it be possible to estimate definite "construct saturations," because no factor corresponding closely to the construct will be available. One can only hope to set upper and lower bounds to the "loading." If "creativity" is defined as something independent of knowledge, then a correlation of .40 between a presumed test of creativity and a test of arithmetic knowledge would indicate that at least 16 per cent of the reliable test variance is irrelevant to creativity as defined. Laboratory performance on problems such as Maier's "hatrack" would scarcely be an ideal measure of creativity, but it would be somewhat relevant. If its correlation with the test is .60, this permits a tentative estimate of 36 per cent as a lower bound. (The estimate is tentative because the test might overlap with the irrelevant portion of the laboratory measure.) The saturation seems to lie between 36 and 84 per cent; a cumulation of studies would provide better limits.

It should be particularly noted that rejecting the null hypothesis does not finish the job of construct validation (**35**, p. 284). The problem is not to conclude that the test "is valid" for measuring the construct variable. The task is to state as definitely as possible the degree of validity the test is presumed to have.

THE LOGIC OF CONSTRUCT VALIDATION

Construct validation takes place when an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconfirming the claim. The philosophy of science which we believe does most justice to actual scientific practice will now be briefly and dogmatically set forth. Readers interested in further study of the philosophical underpinning are referred to the works by Braithwaite (**6**, especially Chapter III), Carnap (**7**; **8**, pp. 56–69), Pap (**51**), Sellars (**55**, **56**), Feigl (**19**, **20**), Beck (**4**), Kneale (**37**, pp. 92–110), Hempel (**29**; **30**, Sec. 7).

The Nomological Net

The fundamental principles are these:

1. Scientifically speaking, to "make clear what something *is*" means to set forth the laws in which it occurs. We shall refer to the interlocking system of laws which constitute a theory as a *nomological network.*

2. The laws in a nomological network may relate (*a*) observable properties or quantities to each other; or (*b*) theoretical constructs to observables; or (*c*) different theoretical constructs to one another. These "laws" may be statistical or deterministic.

3. A necessary condition for a construct to be scientifically admissible is that it occur in a nomological net, at least *some* of whose laws involve observables. Admissible constructs may be remote from observation, i.e., a long derivation may intervene between the nomologicals which implicitly define the construct, and the (derived) nomologicals of type *a*. These latter propositions permit predictions about events. The construct is not "reduced" to the observations, but only combined with other constructs in the net to make predictions about observables.

4. "Learning more about" a theoretical construct is a matter of elaborating the nomological network in which it occurs, or of increasing the definiteness of the components. At least in the early history of a construct the network will be limited, and the construct will as yet have few connections.

5. An enrichment of the net such as adding a construct or a relation to theory is justified if it generates nomologicals that are confirmed by observation or if it reduces the number of nomologicals required to predict the same observations. When observations will not fit into the network as it stands, the scientist has a certain freedom in selecting where to modify the network. That is, there may be alternative constructs or ways of organizing the net which for the time being are equally defensible.

6. We can say that "operations"

which are qualitatively very different "overlap" or "measure the same thing" if their positions in the nomological net tie them to the same construct variable. Our confidence in this identification depends upon the amount of inductive support we have for the regions of the net involved. It is not necessary that a direct observational comparison of the two operations be made—we may be content with an intranetwork proof indicating that the two operations yield estimates of the same network-defined quantity. Thus, physicists are content to speak of the "temperature" of the sun and the "temperature" of a gas at room temperature even though the test operations are nonoverlapping because this identification makes theoretical sense.

With these statements of scientific methodology in mind, we return to the specific problem of construct validity as applied to psychological tests. The preceding guide rules should reassure the "toughminded," who fear that allowing construct validation opens the door to nonconfirmable test claims. *The answer is that unless the network makes contact with observations, and exhibits explicit, public steps of inference, construct validation cannot be claimed.* An admissible psychological construct must be behavior-relevant (**59**, p. 15). For most tests intended to measure constructs, adequate criteria do not exist. This being the case, many such tests have been left unvalidated, or a finespun network of rationalizations has been offered as if it were validation. Rationalization is not construct validation. One who claims that his test reflects a construct cannot maintain his claim in the face of recurrent negative results because these results show that his construct is too loosely defined to yield verifiable inferences.

A rigorous (though perhaps probabilistic) chain of inference is required to establish a test as a measure of a construct. To validate a claim that a test measures a construct, a nomological net surrounding the concept must exist. When a construct is fairly new, there may be few specifiable associations by which to pin down the concept. As research proceeds, the construct sends out roots in many directions, which attach it to more and more facts or other constructs. Thus the electron has more accepted properties than the neutrino; *numerical ability* has more than *the second space factor.*

"Acceptance," which was critical in criterion-oriented and content validities, has now appeared in construct validity. Unless substantially the same nomological net is accepted by the several users of the construct, public validation is impossible. If A uses *aggressiveness* to mean overt assault on others, and B's usage includes repressed hostile reactions, evidence which convinces B that a test measures *aggressiveness* convinces A that the test does not. Hence, the investigator who proposes to establish a test as a measure of a construct must specify his network or theory sufficiently clearly that others can accept or reject it (cf. **41**, p. 406). A consumer of the test who rejects the author's theory cannot accept the author's validation. He must validate the test for himself, if he wishes to show that it represents the construct as *he* defines it.

Two general qualifications are in order with reference to the methodological principles 1–6 set forth at the beginning of this section. Both of them concern the amount of "theory," in any high-level sense of that word, which enters into a construct-defining network of laws or lawlike statements. We do not wish to convey the impression that one always has a very elaborate theoretical network, rich in hypothetical processes or entities.

Constructs as inductive summaries. In the early stages of development of a construct or even at more advanced stages when our orientation is thoroughly practical, little or no theory in the usual sense of the word need be involved. In the extreme case the hypothesized laws are formulated entirely in terms of descriptive (observational) dimensions although not all of the relevant observations have actually been made.

The hypothesized network "goes beyond the data" only in the limited sense that it purports to *characterize* the behavior facets which belong to an observable but as yet only partially sampled cluster; hence, it generates predictions about hitherto unsampled regions of the phenotypic space. Even though no unobservables or high-order theoretical constructs are introduced, an element of inductive extrapolation appears in the claim that a cluster including some elements not-yet-observed has been identified. Since, as in any sorting or abstracting task involving a finite set of complex elements, several nonequivalent bases of categorization are available, the investigator may choose a hypothesis which generates erroneous predictions. The failure of a supposed, hitherto untried, member of the cluster to behave in the manner said to be characteristic of the group, or the finding that a nonmember of the postulated cluster does behave in this manner, may modify greatly our tentative construct.

For example, one might build an intelligence test on the basis of his background notions of "intellect," including vocabulary, arithmetic calculation, general information, similarities, two-point threshold, reaction time, and line bisection as subtests. The first four of these correlate, and he extracts a huge first factor. This becomes a second approximation of the intelligence construct, described by its pattern of loadings on the four tests. The other three tests have negligible loading on any common factor. On this evidence the investigator reinterprets intelligence as "manipulation of words." Subsequently it is discovered that test-stupid people are rated as unable to express their ideas, are easily taken in by fallacious arguments, and misread complex directions. These data support the "linguistic" definition of intelligence and the test's claim of validity *for* that construct. But then a block design test with pantomime instructions is found to be strongly saturated with the first factor. Immediately the purely "linguistic" interpretation of Factor I becomes suspect. This finding, taken together with our initial acceptance of the others as relevant to the background concept of intelligence, forces us to reinterpret the concept once again.

If we simply *list* the tests or traits which have been shown to be saturated with the "factor" or which belong to the cluster, no construct is employed. As soon as we even *summarize the properties* of this group of indicators—we are already making some guesses. Intensional characterization of a domain is hazardous since it selects (abstracts) properties and implies that new tests sharing those properties will behave as do the known tests in the cluster, and that tests not sharing them will not.

The difficulties in merely "characterizing the surface cluster" are strikingly exhibited by the use of certain special and extreme groups for purposes of construct validation. The P_a scale of MMPI was originally derived and cross-validated upon hospitalized patients diagnosed "Psychopathic personality, asocial and amoral type" (**42**). Further research shows the scale to have a limited degree of predictive and concurrent validity for "delinquency" more broadly defined (**5**, **28**). Several studies show associations between P_d and very special "criterion" groups which it would be ludicrous to identify as "*the* criterion" in the traditional sense. If one lists these heterogeneous groups and tries to characterize them intensionally, he faces enormous conceptual difficulties. For example, a recent survey of hunting accidents in Minnesota showed that hunters who had "carelessly" shot someone were significantly elevated on P_d when compared with other hunters (**48**). This is in line with one's theoretical expectations; when you ask MMPI "experts" to predict for such a group they invariably predict P_d or M_a or both. The finding seems therefore to lend some slight support to the construct validity of the P_d scale. But of course it would be nonsense to *define* the P_d component "operationally" in terms of, say, accident proneness. We might try to subsume the original phenotype and the hunting-accident proneness under some broader category, such as "Disposition to violate society's rules, whether legal, moral, or just *sensible*." But now we have ceased to have a neat operational criterion, and are using instead a rather vague and wide-range class. Besides, there is worse to come. We want the class specification to cover a group trend that (nondelinquent) high school students judged by their peer group as least "responsible" score over a full sigma higher on P_d than those judged most "responsible" (**23**, p. 75). Most of the behaviors contributing to such sociometric choices fall well within the range of socially permissible action; the proffered criterion specification is still too restrictive. Again, any clinician familiar with MMPI lore would predict an elevated P_d on a sample of (nondelinquent) professional actors. Chyatte's confirmation of this prediction (**10**) tends to support *both:* (*a*) the theory sketch of "what the P_d factor is, psychologically"; and (*b*) the claim of the P_d scale to construct validity for this hypothetical factor. Let the reader try his hand at writing a brief phenotypic criterion specification that will cover both trigger-happy hunters and Broadway actors! And if he should be ingenious enough to achieve this, does his definition also encompass Hovey's report that high P_d predicts the judgments "not shy" and "unafraid of mental patients" made upon nurses by their supervisors (**32**, p. 143)? And then we have Gough's report that *low* P_d is associated with ratings as "good-natured" (**24**, p. 40), and Roessell's data showing that high P_d is predictive of "dropping out of high school" (**54**). The point is that all seven of these "criterion" dispositions would be readily guessed by any clinician having even superficial familiarity with MMPI interpretation; but to mediate these inferences explicitly requires quite a few hypotheses about dynamics, constituting an admittedly sketchy (but far from vacuous) network defining the genotype *psychopathic deviate.*

Vagueness of present psychological laws. This line of thought leads directly to our second important qualification upon the network schema. The idealized picture is one of a tidy set of postulates which jointly entail the desired theorems; since some of the theorems are coordinated to the observation base, the system constitutes an implicit definition of the theoretical primitives and gives them an indirect empirical meaning. In practice, of course, even the most advanced physical sciences only approximate this ideal. Questions of "categoricalness" and the like, such as logicians raise about pure calculi, are hardly even statable for empirical networks. (What, for example, would be the desiderata of a "well-formed formula" in molar behavior theory?) Psychology works with crude, half-explicit formulations. We do not worry about such advanced formal questions as "whether all molar-behavior statements are decidable by appeal to the postulates" because we know that no existing theoretical network suffices to predict even the *known* descriptive laws. Nevertheless, the sketch of a network is there; if it were not, we would not be saying *anything* intelligible about our constructs. We do not have the rigorous implicit definitions of formal calculi (which still, be it noted, usually permit of a multiplicity of interpretations). Yet the vague, avowedly incomplete network still gives the constructs whatever meaning they do have. When the network is very incomplete, having many strands missing entirely and some constructs tied in only by tenuous threads, then the "implicit definition" of these constructs is disturbingly loose; one might say that the meaning of the constructs is underdetermined. *Since the meaning of theoretical constructs is set forth by stating the laws in which they occur, our incomplete knowledge of the laws of nature produces a vagueness in our constructs* (see Hempel, **30**; Kaplan, **34**; Pap, **51**). We will be able to say "what anxiety is" when we know all of the laws involving it; meanwhile, since we are in the process of discovering these laws, we do not yet know precisely what anxiety is.

CONCLUSIONS REGARDING THE NETWORK AFTER EXPERIMENTATION

The proposition that x per cent of test variance is accounted for by the construct is inserted into the accepted network. The network then generates a testable prediction about the relation of the test scores to certain other variables, and the investigator gathers data. If prediction and result are in harmony, he can retain his belief that the test measures the construct. The construct is at best adopted, never demonstrated to be "correct."

We do not first "prove" the theory, and then validate the test, nor conversely. In any probable inductive type of inference from a pattern of observations, we examine the relation between the total network of theory and observations. The system involves propositions relating test to construct, construct to other constructs, and finally relating some of these constructs to observables. In ongoing research the chain of inference is very complicated. Kelly and Fiske (**36**, p. 124) give a complex diagram showing the numerous inferences required in validating a prediction from assessment techniques, where theories about the criterion situation are as integral a part of the prediction as are the test data. A predicted empirical relationship permits us to test all the propositions leading to that prediction. Traditionally the proposition claiming to interpret the test has been set apart as the hypothesis being tested, but actually the evidence is significant for all parts of the chain. If the prediction is not confirmed, any link in the chain may be wrong.

A theoretical network can be divided into subtheories used in making particular predictions. All the events successfully predicted through a subtheory are of course evidence in favor of that theory. Such a subtheory may be so well confirmed by voluminous and diverse evidence that we can reasonably view a particular experiment as relevant only to the test's validity. If the theory, combined with a proposed test interpretation, mispredicts in this case, it is the latter which must be abandoned. On the other hand, the accumulated evidence for a test's construct validity may be so strong that an instance of misprediction will force us to modify the subtheory employing the construct rather than deny the claim that the test measures the construct.

Most cases in psychology today lie somewhere between these extremes. Thus, suppose we fail to find a greater incidence of "homosexual signs" in the Rorschach records of paranoid patients. Which is more strongly disconfirmed—the Rorschach signs or the orthodox theory of paranoia? The negative finding shows the bridge between the two to be undependable, but this is all we can say. The bridge cannot be used unless one end is placed on solider ground. The investigator must decide which end it is best to relocate.

Numerous successful predictions dealing with phenotypically diverse "criteria" give greater weight to the claim of construct validity than do fewer predictions, or predictions involving very similar behaviors. In arriving at diverse predictions, the hypothesis of test validity is connected each time to a subnetwork largely independent of the portion previously used. Success of these derivations testifies to the inductive power of the test-validity statement, and renders it unlikely that an equally effective alternative can be offered.

Implications of Negative Evidence

The investigator whose prediction and data are discordant must make strategic decisions. His result can be interpreted in three ways:

1. The test does not measure the construct variable.
2. The theoretical network which generated the hypothesis is incorrect.
3. The experimental design failed to test the hypothesis properly. (Strictly speaking this may be analyzed as a special case of 2, but in practice the distinction is worth making.)

For further research. If a specific fault of procedure makes the third a reasonable possibility, his proper response is to perform an adequate study, meanwhile making no report. When faced with the other two alternatives, he may decide that his test does not measure the construct adequately. Following that decision, he will perhaps prepare and validate a new test. Any rescoring or new interpretative procedure for the original instrument, like a new test, requires validation *by means of a fresh body of data*.

The investigator may regard interpretation 2 as more likely to lead to eventual advances. It is legitimate for the investigator to call the network defining the construct into question, if he has confidence in the test. Should the investigator decide that some step in the network is unsound, he may be able to invent an alternative network. Perhaps he modifies the network by splitting a concept into two or more portions, e.g., by designating types of *anxiety*, or perhaps he specifies added conditions under which a generalization holds. When an investigator modifies the theory in such a manner, he is now required to *gather a fresh body of data* to test the altered hypotheses. This step should normally precede publication of the modified theory. If the new data are consistent with the modified network, he is free from the fear that his nomologicals were gerrymandered to fit the peculiarities of his first sample of observations. He can now trust his test to some extent, because his test results behave as predicted.

The choice among alternatives, like any strategic decision, is a gamble as to which course of action is the best investment of effort. Is it wise to modify the theory? That depends on how well the system is confirmed by prior data, and how well the modifications fit available observations. Is it worth while to modify the test in the hope that it will fit the construct? That depends on how much evidence there is—apart from this abortive experiment—to support the hope, and also on how much it is worth to the investigator's ego to salvage the test. The choice among alternatives is a matter of research planning.

For practical use of the test. The consumer can accept a test as a measure of a construct only when there is a strong positive fit between predictions and subsequent data. When the evidence from a proper investigation of a published test is essentially negative, it should be reported as a stop sign to discourage use of the test pending a reconciliation of test and construct, or final abandonment of the test. If the test has not been published, it should be restricted to research use until some degree of validity is established (**1**). The consumer can await the results of the investigator's gamble with confidence that proper application of the scientific method will ultimately tell whether the test has value. Until the evidence is in, he has no justification for employing the test as a basis for terminal decisions. The test may serve, at best, only as a source of suggestions about individuals to be confirmed by other evidence (**15, 47**).

There are two perspectives in test validation. From the viewpoint of the psychological practitioner, the burden of proof is on the test. A test should not be used to measure a trait until its proponent establishes that predictions made from such measures are consistent with the best available theory of the trait. In the view of the test developer, however, both the test and the theory are under scrutiny. He is free to say *to himself privately*, "If my test disagrees with the theory, so much the worse for the theory." This way lies delusion, unless he continues his research using a better theory.

Reporting of Positive Results

The test developer who finds positive correspondence between his proposed interpretation and data is expected to report the basis for his validity claim. Defending a claim of construct validity is a major task, not to be satisfied by a discourse without data. The *Technical Recommendations* have little to say on reporting of construct validity. Indeed, the only detailed suggestions under that heading refer to correlations of the test with other measures, together with a cross reference to some other sections of the report. The two key principles, however, call for the most comprehensive type of reporting. The manual for any test "should report all available information which will assist the user in determining what psychological attributes account for variance in test scores" (**59**, p. 27). And, "The manual for a test which is used primarily to assess postulated attributes of the individual should outline the theory on which the test is based and organize whatever partial validity data there are to show in what way they support the theory" (**59**, p. 28). It is recognized, by a classification as "very desirable" rather than "essential," that the latter recommendation goes beyond present practice of test authors.

The proper goals in reporting construct validation are to make clear (*a*) what interpretation is proposed, (*b*) how adequately the writer believes this interpretation is substantiated, and (*c*) what evidence and reasoning lead him to this belief. Without *a* the construct validity of the test is of no use to the consumer. Without *b* the consumer must carry the entire burden of evaluating the test research. Without *c* the consumer or reviewer is being asked to take *a* and *b* on faith. The test manual cannot always present an exhaustive statement on these points, but it should summarize and indicate where complete statements may be found.

To specify the interpretation, the writer must state what construct he has in mind, and what meaning he gives to that construct. For a construct which has a short history and has built up few connotations, it will be fairly easy to indicate the presumed properties of the construct, i.e., the nomologicals in which it appears. For a construct with a longer history, a summary of properties and references to previous theoretical discussions may be appropriate. It is especially critical to distinguish proposed interpretations from other meanings previously given the same construct. The validator faces no small task; he must somehow communicate a theory to his reader.

To evaluate his evidence calls for a statement like the conclusions from a program of research, noting what is well substantiated and what alternative interpretations have been considered and rejected. The writer must note what portions of his proposed interpretation are speculations, extrapolations, or conclusions from insufficient data. The author has an ethical responsibility to prevent unsubstantiated interpretations from appearing as truths. A claim is unsubstantiated unless the evidence for the claim is public, so that other scientists may review the evidence, criticize the conclusions, and offer alternative interpretations.

The report of evidence in a test manual must be as complete as any research report, except where adequate public reports can be cited. Reference to something "observed by the writer in many clinical cases" is worthless as evidence. Full case reports, on the other hand, may be a valuable source of evidence so long as these cases are representative and negative instances receive due attention. The report of evidence must be interpreted with reference to the theoretical network in such a manner that the reader sees why the author regards a particular correlation or experiment as confirming (or throwing doubt upon) the proposed interpretation. Evidence collected by others must be taken fairly into account.

VALIDATION OF A COMPLEX TEST "AS A WHOLE"

Special questions must be considered when we are investigating the validity of a test which is aimed to provide information about several constructs. In one sense, it is naive to inquire "Is this test valid?" One does not validate a test, but only a principle for making inferences. If a test yields many different types of inferences, some of them can be valid and others invalid (cf. Technical Recommendation C2: "The manual should report the validity of each type of inference for which a test is recommended"). From this point of view, every topic sentence in the typical book on Rorschach interpretation presents a hypothesis requiring validation, and one should validate inferences about each aspect of the personality separately and in turn, just as he would want information on the validity (concurrent or predictive) for each scale of MMPI.

There is, however, another defensible point of view. If a test is purely empirical, based strictly on observed connections between response to an item and some criterion, then of course the validity of one scoring key for the test does not make validation for its other scoring keys any less necessary. But a test may be developed on the basis of a theory which in itself provides a linkage between the various keys and the various criteria. Thus, while Strong's Vocational Interest Blank is developed empirically, it also rests on a "theory" that a youth can be expected to be satisfied in an occupation if he has interests common to men now happy in the occupation. When Strong finds that those with high Engineering interest scores in college are preponderantly in engineering careers 19 years later, he has partly validated the proposed use of the Engineer score (predictive validity). Since the evidence is consistent with the theory on which all the test keys were built, this evidence alone increases the presumption that the *other* keys have predictive validity. How strong is this presumption? Not very, from the viewpoint of the traditional skepticism of science. Engineering interests may stabilize early, while interests in art or management or social work are still unstable. A claim cannot be made that the whole Strong approach is valid just because one score shows predictive validity. But if thirty interest scores were investigated longitudinally and all of them showed the type of validity predicted by Strong's theory, we would indeed be caviling to say that this evidence gives no confidence in the long-range validity of the thirty-first score.

Confidence in a theory is increased as more relevant evidence confirms it, but it is always possible that tomorrow's investigation will render the theory obsolete. The Technical Recommendations suggest a rule of reason, and ask for evidence for each *type* of inference for which a test is recommended. It is stated that no test developer can present predictive validities for all possible criteria; similarly, no developer can run all possible experimental tests of his proposed interpretation. But the recommendation is more subtle than advice that a lot of validation is better than a little.

Consider the Rorschach test. It is used for many inferences, made by means of nomological networks at several levels. At a low level are the simple unrationalized correspondences presumed to exist between certain signs and psychiatric diagnoses. Validating such a sign does nothing to substantiate Rorschach theory. For other Rorschach formulas an explicit a priori rationale exists (for instance, high $F\%$ interpreted as implying rigid control of impulses). Each time such a sign shows correspondence with criteria, its rationale is supported just a little. At a still higher level of abstraction, a considerable body of theory surrounds the general area of *outer control*, interlacing many different constructs. As evidence cumulates, one should be able to decide what specific inference-making chains within this system can be depended upon. One should also be able to conclude—or deny—that so much of the system has stood up under test that one has some confidence in even the untested lines in the network.

In addition to relatively delimited nomological networks surrounding *control* or *aspiration*, the Rorschach interpreter usually has an overriding theory of the test as a whole. This may be a psychoanalytic theory, a theory of perception and set, or a theory stated in terms of learned habit patterns. Whatever the theory of the interpreter, whenever he validates an inference from the system, he obtains some reason for added confidence in his overriding system. His total theory is not tested, however, by experiments dealing with only one limited set of constructs. The test developer must investigate far-separated, independent sections of the network. The more diversified the predictions the system is required to make, the greater confidence we can have that only minor parts of the system will later prove faulty. Here we begin to glimpse a logic to defend the judgment that the test and its whole interpretative system is valid at some level of confidence.

There are enthusiasts who would conclude from the foregoing paragraphs that since there is some evidence of correct, diverse predictions made from the Rorschach, the test as a whole can now be accepted as validated. This conclusion overlooks the negative evidence. Just one finding contrary to expectation, based on sound research, is sufficient to wash a whole theoretical structure away. Perhaps the remains can be salvaged to form a new structure. But this structure now must be exposed to fresh risks, and sound negative evidence will destroy it in turn. There is sufficient negative evidence to prevent acceptance of the Rorschach and its accompanying interpretative structures as a whole. So long as any aspects of the overriding theory stated for the test have been disconfirmed, this structure must be rebuilt.

Talk of areas and structures may seem not to recognize those who would interpret the personality "globally." They may argue that a test is best validated in matching studies. Without going into detailed questions of matching methodology, we can ask whether such a study validates the nomological network "as a whole." The judge does employ some network in arriving at his conception of his subject, integrating specific inferences from specific data. Matching studies, if successful, demonstrate only that each judge's interpretative theory has some validity, that it is not completely a fantasy. Very high consistency between judges is required to show that they are using the same network, and very high success in matching is required to show that the network is dependable.

If inference is less than perfectly dependable, we must know which aspects of the interpretative network are least dependable and which are most dependable. Thus, even if one has considerable confidence in a test "as a whole" because of frequent successful inferences, one still returns as an ultimate aim to the request of the Technical Recommendation for separate evidence on the validity of each type of inference to be made.

RECAPITULATION

Construct validation was introduced in order to specify types of research required in developing tests for which the conventional views on validation are inappropriate. Personality tests, and some tests of ability, are interpreted in terms of attributes for which there is no adequate criterion. This paper indicates what sorts of evidence can substantiate such an interpretation, and how such evidence is to be interpreted. The following points made in the discussion are particularly significant.

1. A construct is defined implicitly by a network of associations or propo-

sitions in which it occurs. Constructs employed at different stages of research vary in definiteness.

2. Construct validation is possible only when some of the statements in the network lead to predicted relations among observables. While some observables may be regarded as "criteria," the construct validity of the criteria themselves is regarded as under investigation.

3. The network defining the construct, and the derivation leading to the predicted observation, must be reasonably explicit so that validating evidence may be properly interpreted.

4. Many types of evidence are relevant to construct validity, including content validity, interitem correlations, intertest correlations, test-"criterion" correlations, studies of stability over time, and stability under experimental intervention. High correlations and high stability may constitute either favorable or unfavorable evidence for the proposed interpretation, depending on the theory surrounding the construct.

5. When a predicted relation fails to occur, the fault may lie in the proposed interpretation of the test or in the network. Altering the network so that it can cope with the new observations is, in effect, redefining the construct. Any such new interpretation of the test must be validated by a fresh body of data before being advanced publicly. Great care is required to avoid substituting a posteriori rationalizations for proper validation.

6. Construct validity cannot generally be expressed in the form of a single simple coefficient. The data often permit one to establish upper and lower bounds for the proportion of test variance which can be attributed to the construct. The integration of diverse data into a proper interpretation cannot be an entirely quantitative process.

7. Constructs may vary in nature from those very close to "pure description" (involving little more than extrapolation of relations among observation-variables) to highly theoretical constructs involving hypothesized entities and processes, or making identifications with constructs of other sciences.

8. The investigation of a test's construct validity is not essentially different from the general scientific procedures for developing and confirming theories.

Without in the least *advocating* construct validity as preferable to the other three kinds (concurrent, predictive, content), we do believe it imperative that psychologists make a place for it in their methodological thinking, so that its rationale, its scientific legitimacy, and its dangers may become explicit and familiar. This would be preferable to the widespread current tendency to engage in what actually amounts to construct validation research and use of constructs in practical testing, while talking an "operational" methodology which, if adopted, would force research into a mold it does not fit.

REFERENCES

1. AMERICAN PSYCHOLOGICAL ASSOCIATION. *Ethical standards of psychologists.* Washington, D.C.: American Psychological Association, Inc., 1953.

2. ANASTASI, ANNE. The concept of validity in the interpretation of test scores. *Educ. psychol. Measmt,* 1950, **10**, 67–78.

3. BECHTOLDT, H. P. Selection. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1237–1267.

4. BECK, L. W. Constructions and inferred entities. *Phil. Sci.,* 1950, **17**. Reprinted in H. Feigl and M. Brodbeck

(Eds.), *Readings in the philosophy of science.* New York: Appleton-Century-Crofts, 1953. Pp. 368–381.

5. Blair, W. R. N. A comparative study of disciplinary offenders and non-offenders in the Canadian Army. *Canad. J. Psychol.*, 1950, **4**, 49–62.
6. Braithwaite, R. B. *Scientific explanation.* Cambridge: Cambridge Univer. Press, 1953.
7. Carnap, R. Empiricism, semantics, and ontology. *Rév. int. de Phil.*, 1950, **II**, 20–40. Reprinted in P. P. Wiener (Ed.), *Readings in philosophy of science*, New York: Scribner's, 1953. Pp. 509–521.
8. Carnap, R. *Foundations of logic and mathematics. International encyclopedia of unified science*, **I**, No. 3. Pages 56–69 reprinted as "The interpretation of physics" in H. Feigl and M. Brodbeck (Eds.), *Readings in the philosophy of science.* New York: Appleton-Century-Crofts, 1953. Pp. 309–318.
9. Child, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–171.
10. Chyatte, C. Psychological characteristics of a group of professional actors. *Occupations*, 1949, **27**, 245–250.
11. Cronbach, L. J. *Essentials of psychological testing.* New York: Harper, 1949.
12. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
13. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–335.
14. Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
15. Cronbach, L. J. The counselor's problems from the perspective of communication theory. In Vivian H. Hewer (Ed.), *New perspectives in counseling.* Minneapolis: Univer. of Minnesota Press, 1955.
16. Cureton, E. E. Validity. In E. F. Lindquist (Ed.), *Educational measurement.* Washington, D. C.: American Council on Education, 1950. Pp. 621–695.
17. Damrin, Dora E. A comparative study of information derived from a diagnostic problem-solving test by logical and factorial methods of scoring. Unpublished doctor's dissertation, Univer. of Illinois, 1952.
18. Eysenck, H. J. Criterion analysis—an application of the hypothetico-deductive method in factor analysis. *Psychol. Rev.*, 1950, **57**, 38–53.
19. Feigl, H. Existential hypotheses. *Phil. Sci.*, 1950, **17**, 35–62.
20. Feigl, H. Confirmability and confirmation. *Rév. int. de Phil.*, 1951, **5**, 1–12. Reprinted in P. P. Wiener (Ed.), *Readings in philosophy of science.* New York: Scribner's, 1953. Pp. 522–530.
21. Gaylord, R. H. Conceptual consistency and criterion equivalence: a dual approach to criterion analysis. Unpublished manuscript (PRB Research Note No. 17). Copies obtainable from ASTIA-DSC, AD-21 440.
22. Goodenough, Florence L. *Mental testing.* New York: Rinehart, 1950.
23. Gough, H. G., McClosky, H., & Meehl, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, **47**, 73–80.
24. Gough, H. G., McKee, M. G., & Yandell, R. J. Adjective check list analyses of a number of selected psychometric and assessment variables. Unpublished manuscript. Berkeley:IPAR, 1953.
25. Guilford, J. P. New standards for test evaluation. *Educ. psychol. Measmt.* 1946, **6**, 427–439.
26. Guilford, J. P. Factor analysis in a test-development program. *Psychol. Rev.*, 1948, **55**, 79–94.
27. Gulliksen, H. Intrinsic validity. *Amer. Psychologist*, 1950, **5**, 511–517.
28. Hathaway, S. R., & Monachesi, E. D. *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. of Minnesota Press, 1953.
29. Hempel, C. G. Problems and changes in the empiricist criterion of meaning. *Rév. int. de Phil.*, 1950, **4**, 41–63. Reprinted in L. Linsky, *Semantics and the philosophy of language.* Urbana: Univer. of Illinois Press, 1952. Pp. 163–185.
30. Hempel, C. G. *Fundamentals of concept formation in empirical science.* Chicago: Univer. of Chicago Press, 1952.
31. Horst, P. The prediction of personal adjustment. *Soc. Sci. Res. Council Bull.*, 1941, No. 48.
32. Hovey, H. B. MMPI profiles and personality characteristics. *J. consult. Psychol.*, 1953, **17**, 142–146.
33. Jenkins, J. G. Validity for what? *J. consult. Psychol.*, 1946, **10**, 93–98.
34. Kaplan, A. Definition and specification of meaning. *J. Phil.*, 1946, **43**, 281–288.
35. Kelly, E. L. Theory and techniques of assessment. *Annu. Rev. Psychol.*, 1954, **5**, 281–311.
36. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psy-*

chology. Ann Arbor: Univer. of Michigan Press, 1951.

37. KNEALE, W. *Probability and induction.* Oxford: Clarendon Press, 1949. Pages 92–110 reprinted as "Induction, explanation, and transcendent hypotheses" in H. Feigl and M. Brodbeck (Eds.), *Readings in the philosophy of science.* New York: Appleton-Century-Crofts, 1953. Pp. 353–367.
38. LINDQUIST, E. F. *Educational measurement.* Washington, D. C.: American Council on Education, 1950.
39. LUCAS, C. M. Analysis of the relative movement test by a method of individual interviews. *Bur. Naval Personnel Res. Rep.*, Contract Nonr-694 (00), NR 151-13, Educational Testing Service, March 1953.
40. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
41. MACFARLANE, JEAN W. Problems of validation inherent in projective methods. *Amer. J. Orthopsychiat.*, 1942, **12**, 405–410.
42. MCKINLEY, J. C., & HATHAWAY, S. R. The MMPI: V. Hysteria, hypomania, and psychopathic deviate. *J. appl. Psychol.*, 1944, **28**, 153–174.
43. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The K scale. *J. consult. Psychol.*, 1948, **12**, 20–31.
44. MEEHL, P. E. A simple algebraic development of Horst's suppressor variables. *Amer. J. Psychol.*, 1945, **58**, 550–554.
45. MEEHL, P. E. An investigation of a general normality or control factor in personality testing. *Psychol. Monogr.*, 1945, **59**, No. 4 (Whole No. 274).
46. MEEHL, P. E. *Clinical vs. statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954.
47. MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
48. *Minnesota Hunter Casualty Study.* St. Paul: Jacob Schmidt Brewing Company, 1954.
49. MOSIER, C. I. A critical examination of the concepts of face validity. *Educ. psychol. Measmt*, 1947, **7**, 191–205.
50. MOSIER, C. I. Problems and designs of cross-validation. *Educ. psychol. Measmt*, 1951, **11**, 5–12.
51. PAP, A. Reduction-sentences and open concepts. *Methodos*, 1953, **5**, 3–30.
52. PEAK HELEN. Problems of objective observation. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden Press, 1953. Pp. 243–300.
53. PORTEUS, S. D. *The Porteus maze test and intelligence.* Palo Alto: Pacific Books, 1950.
54. ROESSEL, F. P. MMPI results for high school drop-outs and graduates. Unpublished doctor's dissertation, Univer. of Minnesota, 1954.
55. SELLARS, W. S. Concepts as involving laws and inconceivable without them. *Phil. Sci.*, 1948, **15**, 287–315.
56. SELLARS, W. S. Some reflections on language games. *Phil. Sci.*, 1954, **21**, 204–228.
57. SPIKER, C. C., & MCCANDLESS, B. R. The concept of intelligence and the philosophy of science. *Psychol. Rev.*, 1954, **61**, 255–267.
58. Technical recommendations for psychological tests and diagnostic techniques: preliminary proposal. *Amer. Psychologist*, 1952, **7**, 461–476.
59. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull. Supplement*, 1954, **51**, 2, Part 2, 1–38.
60. THURSTONE, L. L. The criterion problem in personality research. *Psychometric Lab. Rep.*, No. 78. Chicago: Univer. of Chicago, 1952.

Received for early publication February 18, 1955

Psychological Bulletin
Vol. 52, No. 4, 1955

PSYCHOLOGY IN THE ARAB NEAR EAST

E. TERRY PROTHRO AND LEVON H. MELIKIAN
American University of Beirut

It is the purpose of this paper to present a general picture of psychology in the Arab Near East. Particular attention will be given to places and institutions where most of the psychological activity is concentrated, the type of work carried on in each of those places, and persons who seem to be leaders in the field. The emphasis is on the contemporary scene, but background factors are introduced where they seem to shed light on the current situation.

There is some disagreement among scholars on the inclusiveness of the terms "Near East" and "Middle East." We have included Egypt, Lebanon, Syria, Iraq, Jordan, Saudi Arabia, Bahrein Island, and Kuwait under the phrase "Arab Near East." Some British writers use "Arab Middle East" to refer to the same area. The population of this region is approximately 40 million, of which about half is in Egypt. The total area is more than 1,600,000 square miles. Thus the region is about one fourth as populous, and more than half as large, as the continental United States. Population density is considerably greater than these figures would suggest, for well over three-fourths of the area is desert and the population is concentrated in the fertile, watered areas.

The recorded political history of the Near East goes back thousands of years. We shall not attempt to trace the evolvement of these countries to their present form. It is sufficient here to note that all are independent sovereign nations except Bahrein and Kuwait. These two are British protectorates.

The information presented in this paper is based principally on firsthand observation and personal interviews by the authors. The senior author has taught in the Near East for more than two years. The junior author was born in the Near East and has spent most of his life there. Preliminary information about psychology in the area was gathered through correspondence and informal discussions. When it was decided to make an intensive survey of the area, information was solicited by mail from all known psychologists in each country. On the basis of this information, interviews were scheduled in the spring of 1953 with leading psychologists in each country. One or both of us then visited each state, talked at length with the selected psychologists, and personally inspected laboratories, libraries, classrooms, etc. Visits were not made to Kuwait, Bahrein, or Saudi Arabia because we knew of no psychologists there. For information about these countries we relied on the statements of educators from them who visited Lebanon, and on the report of Dr. Habib Kurani, who inspected some of their educational facilities in 1952. Because of the relative importance of Egyptian psychology, additional information about work there was obtained through the kind cooperation of Dr. Hanna Rizk of the American University in Cairo. He prepared and circulated a detailed questionnaire to all Egyptians engaged in full or part-time psychological work. From the returned questionnaires we were able to get an unusually complete picture of psychology in Egypt.

GENERAL CHARACTERISTICS

Psychology in the Arab Near East sprang from and is nourished by Western psychology. British, French, and American influences can be seen—and distinguished—in the work of psychologists throughout the area. The kind and strength of these influences is determined by several factors: the nature of Arab culture, the social and political forces exerted by Western nations, and the comparative vigor of the several transplanted psychologies. Consequently Near Eastern psychology is of interest not only for its own sake, but also as a field for the study of forces affecting the development of various psychological interests and viewpoints.

In the days of the Ottoman Empire, the French school system became the model for Arab schools. With the breakup of that empire in World War I, French influence gained in the Levant but schools of the British type began to flourish in Iraq, Jordan, Palestine, and Egypt; today government schools still follow French and British organizational patterns.

In Syria, Lebanon, and Palestine there are many private schools which were established by Western missionaries before the Ottomans created any local schools. The private schools with foreign support continue to exercise great influence on the education of the area. Consequently the Arab who begins his study of psychology in his own land, whether in public or private colleges, will approach the subject from a Western point of view. If he seeks graduate training, he will probably go to France, Britain, or the United States.

There is remarkably little communication among psychologists of the area. Dependence on Western universities for advanced training reduces the amount of contact among Arabs of different nationalities. A French-trained psychologist in Lebanon may not read the journal published by a French-trained psychologist in Egypt, although he will be quite *au courant* of the publications from Paris and Lyon. A British-trained psychologist in Damascus may know nothing of the work of the Baghdad psychologist who also received his doctorate in England. This isolation is augmented by severe travel restrictions imposed by many of the Arab states. Since there is little integration of psychological activities in this area, we shall deal with psychology in each of the several countries in turn. The order of presentation is based on our judgment of the amount of psychological activity to be found in each country.

EGYPT[1]

There are approximately 40 persons in Egypt who are engaged full time in psychological work. More than 30 of these are engaged in academic work, and most of the remainder are employed by the government. Two psychologists were cabinet ministers at the time this survey was made (2).

Two generalizations can be made about psychology in Egypt. First, the influence of British universities is very strong. Second, there is a close alliance between psychology and education. The second fact is probably related to the first. There are many individuals, of course, to whom these statements would not apply, and there is some evidence that British influence is less strong today than five years ago. Nevertheless, more than half of the psychologists in Egypt today received their advanced degrees from British institutions, and British journals rank next to Egyptian journals as vehicles for publication of research data. Only two persons reported having published in American journals, and two reported having published in France.

---

[1] We wish to express our appreciation to Dr. Hanna Rizk of the American University at Cairo for invaluable assistance in collecting data on Egyptian psychologists.

Nearly all of Egypt's psychologists are to be found in Cairo and Alexandria. A majority of the psychologists engaged in academic work are in one or another of the Institutes of Education. The Higher Institute of Education for Men, now a part of Cairo's Ibrahim University, has 11 psychologists on its staff. Four psychologists are with the Higher Institute of Education for Women in Cairo, two are with the Higher Institute of Education in Alexandria, and at least six more are on the staffs of the various normal schools. As might be expected from their institutional affiliations, more psychologists publish in the *Egyptian Journal of Education* than in the *Egyptian Journal of Psychology*. Indeed, this latter publication has recently changed from a quarterly to an annual.

Dean of the Institute of Higher Education for Men, Cairo, is Dr. Abdel Aziz El Koussy, whose graduate training was taken at Birmingham under Valentine and at London under Spearman and Burt. Specialists in clinical and developmental psychology, statistics, tests, and learning are on the Institute's faculty. These specialists hold Ph.D. degrees (or the equivalent) from Cambridge, London, Reading, and Leeds. All undergraduate students at the Institute take at least one course in psychology, and students in the newly organized graduate school take approximately half of their course work in psychology. In 1952, 81 persons were taking graduate work, and 24 had passed the master's examination. Of those who had passed, about one-third were actively working on a master's thesis. The Institute is authorized to award both the M.A. and Ph.D. in educational psychology. At this writing no Ph.D. had been awarded.

The laboratory at the Institute is similar to that in many American universities. Several rooms are available; each is furnished with worktables and cabinets filled with apparatus. The apparatus include chronoscopes, mazes, mirror-drawing devices, galvanometers, kymographs, manual dexterity boards, instruments designed to measure mechanical ability, intelligence tests, and a large variety of paper-and-pencil tests. Undergraduate students who take experimental psychology are expected to perform four experiments from a list of experiments provided by the instructor. Candidates for the master's degree perform 20 experiments. English and American books serve as references. Foreign journals are not readily available to the students.

The Institute offers some work in clinical psychology and there is a clinic connected with it. The demands for the services of the clinic far exceed its capacity. Patients are limited to children and students. Three professors devote full time to the clinic, and a neurologist, a psychiatrist, and a social worker are available on a part-time basis. Dr. Ishak Ramzy is in charge. He holds a Ph.D. from London and was trained in psychoanalysis at the London Institute of Psychoanalysis, where he studied with Anna Freud.

There is an active testing movement in Egypt, and the Institute is one of the foci of this activity. Adaptations of the Binet and of several of the Spearman tests have been in use for years. Pioneers and leaders in this area were Dr. Koussy and Dr. Ismael Kabbani, formerly with the Institute and now Minister of Education. At the present time the Army is using intelligence tests and other aptitude tests.

The Faculty of Arts of Ibrahim University, Cairo, offers a *license* in psychology which is roughly equivalent to the American master of arts degree. Chairman of the department is Dr. Mustafa Ziwer who holds both the M.D. and *Dr. es Let.* from the University of Paris. The latter degree was earned under Piéron. Although the department is small, an imposing array of courses is offered by drawing on the staff of the Institute of Higher Education and on the staff of the Faculty of Arts, University of Cairo. The program requires four years beyond the *baccalaureate deuxieme partie*, and the first class to have completed this entire new program was graduated in 1954. It is expected that the graduates will be employed by the Ministries of Education and Social Welfare. Work toward a Ph.D. degree is to be offered in the future.

At the University of Cairo, the psychology department is a part of the philosophy section of the Faculty of Arts. Students working toward a *license* in philosophy take one psychology course in each of the last three years. Chairman of the department is Dr. Yusuf Mourad, a graduate of the University of Paris. He is co-editor, with Dr. Ziwer, of the *Egyptian Journal of Psychology*, and has contributed to it extensively. For his advanced courses he assigns works in French by Guillaume, Piéron, Wallon, and others, and French translations of such American authors as Woodworth, Carmichael, and Krech and Crutchfield. Courses in experimental psychology are offered at Fouad by Dr. Muhammad Nagati, who did his graduate work at Yale University.

The Higher Institute of Education for Women, Cairo, and the Alexandria Higher Institute of Education for Men both offer work in educational psychology. Each of these institutes has at least three full-time professors in this field. The American University at Cairo offers courses in psychology as part of its education and social studies curricula. Psychological training is also a part of the program given by the Cairo Institute of Social Work.

Al Azhar, Cairo, one of the world's oldest universities, is traditionally devoted primarily to theological training. There are today three sections at the higher level: theology, Islamic law, and Arabic language. The section devoted to the teaching of the Arabic language now offers courses in general and educational psychology, and the theology section offers one course in psychology.

LEBANON[2]

Since the time of the Crusades, and especially for the past hundred years, Lebanon has had close cultural ties with the West. At present one finds in the book stores of its cities large numbers of Western books, including scientific and pseudoscientific works on psychology. Works on psychoanalysis are so popular that a leading Beirut newspaper ran in 1953 a series of editorials denouncing the "efforts of Western psychologists" to turn attention from social ills to problems of infantile frustration. There is even one dianetics practitioner in Beirut who received training from Hubbard himself.

---

[2] We wish to express our appreciation to Miss Suad Khalluf for assistance in obtaining information about psychology in Lebanon.

There are three universities in Lebanon, and these universities are the foci of psychological activity in the country. *The Centre d'Etudes Supérieures Français de Beyrouth* is a branch of the Arts and Sciences Faculty of the University of Lyon. It offers a *license* degree which requires that the student complete five certificates. The five can be obtained in two or three years. Of the many certificates which may be chosen, three are in the field of psychology: general, social, and educational. Each certificate requires from six to eight hours per week for an academic year. The professors of the psychology courses hold chairs of philosophy. They place great emphasis on preparation for the annual examinations, which are issued from Lyon. They report, in familiar terms, that their teaching keeps them too busy for research.

The first psychological work published in the Arabic language in modern times seems to have been *'Al durūs al 'awaliyah fil falsafah al 'qliyah* (First Lessons in Mental Philosophy). It was written in 1874 by the American missionary Daniel Bliss, who was founder and first president of the Syrian Protestant College (now the American University of Beirut). Today the psychology department of the American University resembles that of the liberal arts college in the United States. Courses are taught in English, and American textbooks are used. An eight-hour course in experimental and a three-hour course in statistics are required of all majors. The university emphasizes research, and for more than a decade the department has concentrated on research in social psychology. This tradition was begun by S. C. Dodd, now at the University of Washington, who was for many years chairman of the sociology and psychology departments. Closely allied with the American University is the Beirut College for Women. The College stresses work in developmental, guidance, and educational psychology.

The Lebanese National University held its first classes in 1950, and the psychology program there is in an embryonic stage. Some courses are offered now and it is planned to offer a certificate in psychology in 1956. The organization and content of the courses reflect the influence of the French university and, to a lesser extent, of the American University.

SYRIA

Syrian psychology centers in the Syrian State University located in Damascus. The university is quite new. Although the Schools of Medicine and Law date back to 1923, it was only in 1946 that the Schools of Arts, Science, and Engineering, and the Higher Normal School, were joined with the older faculties (1). Some of the buildings occupied by the university are former Ottoman and French army barracks. Psychology courses are offered in both the philosophy faculty of the College of Arts and Sciences and in the education department of the Higher Normal School. In order to obtain a *license* in the philosophy faculty, a student must study philosophy, psychology, and sociology. If he receives the degree, he is entitled to teach these subjects in secondary schools. The secondary schools, like those of France, carry the students to approximately the American sophomore level. Of the four professors on the philosophy faculty, two are philosophers and the other two are psychologists on loan from the Higher Normal School.

The education department is quite active in psychology, and the members of its faculty offer a wide variety of courses. Chairman of the department is Dr. Fakhir Akil, who received B.A. and M.A. degrees from the American University of Beirut, and a Ph.D. from London. His doctoral work was done under Drs. Philpott and Burt. Each of the other full-time professors holds a *license* from Egypt and is completing a doctorate in France. Consequently, Syrian psychology is a mixture of American, English, and French approaches. The professors are writing textbooks in Arabic, and they assign collateral reading in both French and American works. Woodworth's *Experimental Psychology* and Morgan's *Physiological Psychology* are used in their French editions. The laboratory is well equipped with pursuit meters, mirror-drawing apparatus, form boards, intelligence test materials, personality test materials (including Rorschach, TAT, Spearman Abilities, and Szondi), and many other devices. We were impressed to learn at the time of our visit to the laboratory that the annual budget for new laboratory equipment was approximately $4,500.

Students achieving satisfactory scores on entrance examinations may study at the Higher Normal School at government expense. If they wish to accept the subsidy, they sign a contract with the Ministry of Education whereby they agree to work for a specific number of years for that ministry upon completion of their work. Moreover, the top student each year is sent abroad to work for a doctor's degree. Partly as a result of this arrangement, there is strong student interest in educational psychology. The trend seems to be away from philosophical psychology, in the French tradition, toward educational psychology of the British and American variety. This trend may be a function of Dr. Akil's leadership, and it is too early to say whether it will continue when students now studying abroad return to Syria. Because of currency problems, transportation costs, and other factors, more Syrians are studying in France than in Britain and America together (4). In Syria as in Egypt, there is more psychology work in the education department than in the philosophy faculty.

IRAQ

A visit to modern Baghdad gives one the impression that there is a lively curiosity about psychology in literate circles. Bookstores stock and presumably sell many English-language psychology textbooks, even though these textbooks are not assigned in any Iraqi university. Woodworth's *Experimental Psychology*, for example, is apparently read by some Iraqis out of sheer intellectual curiosity!

Although many of the professors are American educated, education in Iraq is dominated by persons trained in Britain. As in Egypt, educational psychology is taught in both urban and rural teacher training institutions. Teachers of these courses are persons who were trained primarily in the field of education.

There is a department of education and psychology in the Higher Teacher's Training College, located in Baghdad. Several courses are offered in psychology, and there are plans for expansion. The principal members of the staff are Dr. Ibrahim Muhyi, who completed his doctorate at Columbia University, and Dr. Abdul Aziz Al-Bassam, whose graduate study was at University College, London. Each of these professors also offers courses at Baghdad's College of Arts and Sciences and at the Queen Alliya College for Women. This last-named school offers psychology in the department of social work. As far as we know, there is no psychological laboratory in Iraq, and little psychological research.

THE PERSIAN GULF AREA

There is hardly any psychological activity on the West (Arab) side of the Persian Gulf. In Jordan there are two teacher training institutions where Egyptian-trained educators offer elementary work in psychology. There are no trained psychologists in Bahrein or Kuwait, and none in the Arabian peninsula except the Americans who are employed by the Arabia-American Oil Company.

CONCLUSION

From this brief survey it can be seen that psychology in the Arab Near East is essentially academic. Where the French academic tradition prevails, psychology is closely allied with philosophy. Where British academic tradition prevails, psychology is allied with education, and there is a strong emphasis on testing and measurement. There is some concern with research in Egypt and Lebanon, but this activity is subordinate throughout the area to translation of important Western books and adaptation to local use of Western tests.

In his discussion of "the sociology of psychology" Sanford (**3**, p. 85) has suggested that dictatorial regimes are antipsychological. Yet the military regimes in both Egypt and Syria seem eager to utilize whatever contributions psychology can make to the perceived social problems. Interest in psychology on the part of these governments exceeds that of the parliamentary governments which preceded them. This does not demonstrate that *antidemocratic* trends favor psychology, for the political form is not always the best index of the essential nature of a society.

The prognosis is favorable for the growth of psychology in this part of Western Asia. Changes in long-established social institutions, increased urbanization, increased literacy, and rising per capita income all point to increased utilization of psychological understanding. There is some suspicion that psychology is an instrument of Western imperialism, but there is reason to hope that this suspicion will be overcome as increasing numbers of Arabs become psychologists.

APPENDIX

BIBLIOGRAPHY OF SELECTED TITLES

There is no published bibliography of recent books in the Arabic language which deals with psychological topics. The best source of information about such books is probably *'Al Sijil al Thakafi* (Cultural Register), published annually by the Egyptian Ministry of Education in Cairo. The register contains information about publications in many fields, including psychology and education.

We have listed below some books which we consider to be representative of the best works on psychology published in Arabic from 1945 to 1953. The list is representative rather than inclusive.

1. 'Al 'Ahwani, 'Ahmad Fu'ad. [Outline of psychology.] Cairo: Amiriya Press, 1948.
2. Al Athama, Wafic. [Modern psychology.] Damascus: Hashimiyah Press, 1950.
3. Al Draubi, Sami, & 'Al Jamali, Hafiz. [Psychology.] Damascus: Dar-'al-Yaqtha Press, 1949.
4. Al Khawli, William. [The integrated personality.] Cairo: Mukhaymar Press, 1948.
5. Al Qusi, Abdul Aziz. [Foundations of

mental health.] (2nd Ed.) Cairo: Al Nahda Press, 1952.

6. Al Qusi, Abdul Aziz. [Foundations of Psychology.] Cairo: Al Nahda Library, 1950.
7. Al Qusi, Abdul Aziz, *et al.* [Language and thought.] Cairo: Amiriyah Press, 1948.
8. Al Qusi, Abdul Aziz. [Statistics in education and psychology.] Cairo: Al Nahda Library, 1949.
9. Al Mulayhi, Abdul Mun'im. [Psychological development.] Cairo: Misr Library, 1950.
10. Al Tawil, Tawfic. [Dreams.] Cairo: Al Adab Press, 1945.
11. 'Akil, Fakhir. [Psychology and its application to education.] Damascus: Maktabat Al'uloom Wal'Adab, 1945.
12. Barakat, Mohamad Khalifah. [Analysis of personality.] Cairo: Misr Library, 1951.
13. Barakat, Mohamad Khalifah. [Psychological clinics.] Cairo: Misr Library, 1952.
14. Jirjis, Sabri. [The problem of psychopathic behavior.] Cairo: Dar Al Ma'rif, 1952.
15. Khalaf Allah Muhammed. [A psychological approach to the study of literature and literary criticism.] Cairo: Lajnat Al ta'leaf Wal tarjamat Wal Nashr, 1947.
16. Mitri, 'Amin. [The feebleminded.] Alexandria: Dar Nashr 'Al Thaqafah, 1948.
17. Murad, Yusuf. [Elements of general psychology.] Cairo: Dar Al Ma'rif Press, 1948.
18. Najati, Mohamad 'Uthman. [Military psychology.] Cairo: Dar Al Ma'rif, 1953.
19. Ramzi, Ishaq. [Individual psychology.] Cairo: Dar Al Ma'rif Press, 1946.
20. Saliba, Jamil. [Psychology.] Damascus: 'Ibn Zaydoun Press, 1948.
21. Salih, 'Ahmad. [Educational psychology.] Cairo: Al Nahda Library, 1951.
22. Subay'i, 'Adnan. [A survey of child psychology.] Damascus: 'Arafah Library, 1952.

REFERENCES

1. MATTHEWS, R., & AKRAWI, M. *Education in Arab countries of the Near East.* Washington: American Council on Education, 1949.
2. PROTHRO, E. T. Psychologists in high posts in government. *Amer. Psychologist,* 1953, **10**, 597.
3. SANFORD, F. H. Toward a sociology of psychology. *Amer. Psychologist,* 1952, **7**, 83–85.
4. WILLIAMS, H. H. Syrians studying abroad. New York: *Institute of International Education*, 1952.

Received May 15, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 4, 1955

THE ROLE OF MOTIVATION IN VERBAL LEARNING AND PERFORMANCE[1]

I. E. FARBER
State University of Iowa

The factors affecting performance in learning situations are customarily divided into two classes—the associative and the nonassociative. The former have been the subject of most research in the area of verbal behavior, and the information concerning them constitutes by far the better part, in quality as well as quantity, of what is known about this kind of behavior. In recent years, however, there has been an increasing interest in those factors that are commonly supposed to have nonassociative properties, particularly those subsumed under the heading of "motivation."

Although the term "motivational" is often used in reference to nonassociative factors, conventional psychological usage imputes to motives an associative aspect also. Melton, for instance, considers motives as having three functions: the energizing, the selective, and the directing. According to his view, the energizing function "... may be most simply conceptualized as a result of the lowering of the *general* (italics added) reaction threshold by the motive" (**30**, p. 672); the selective function refers to the reinforcing effects of the achievement of goals relevant to the motive; and the directing function is associative or cognitive, considering motives, as part of the stimulus complex, as having a given habit strength in respect to various specific responses. According to Shaffer (**41**), a motive involves both "drive" and "mechanism," the former term denoting a variable that arouses mass activity, and the latter, an acquired response tendency. Hull (**23**), in his discussion of primary motivation, supposed that while " . . . all drives alike are able to sensitize all habits" (p. 247), "associated with every drive (D) is a characteristic drive stimulus (S_D) . . . " (p. 253), which gives specificity and direction to behavior. And Dollard and Miller (**13**), in discussing the functional properties of stimuli, refer to their drive aspect—that which impels action generally, and their cue aspect—that which leads to differential response.

The supposition that motivational variables have steering (associative) properties as well as energizing and reinforcing (drive) properties does not, of course, imply that the two kinds of properties are identical. Indeed, the virtue of the foregoing analyses of the concept of motivation lies in their separation of these functions. The associative function of a motive or any other variable is identified in terms of its tendency to elicit or facilitate a limited class of responses only. The drive function of a variable is demonstrated if: (*a*) its presence energizes or intensifies indiscriminately all reaction tendencies existing in a given situation; and/or (*b*) its elimination or reduction in magnitude is reinforcing, i.e., leads to the increased probability of recurrence, in the same situation, of the responses preceding its modification. Motives, it is presumed, have both functions, and if a variable clearly does not have both, its status as a motive is questionable.

---

[1] The present paper is a revision of a paper entitled "Motivational factors in verbal learning," presented May 8, 1953, in the *Symposium on Psychology of Learning Basic to Military Training Problems*, sponsored by the Panel on Training and Training Devices, Research and Development Board. It was prepared as part of a project on the influence of motivation on performance in learning, under Contract N9 onr-93802, Project NR 154-107, between the State University of Iowa and the Office of Naval Research. The helpful criticisms of this manuscript by Drs. K. W. Spence and J. S. Brown are acknowledged with thanks.

But, as Brown has recently pointed out, "an examination of contemporary discussions of motivation suggests that one of the major sources of misunderstanding is the failure of most writers to distinguish clearly between drives or motives, on the one hand, and habits or reaction tendencies on the other" (**5**, p. 2). According to his view, all instances of directed behavior are attributable, not to drives or motives as such, but rather to the capacity of stimulus cues, including those accompanying drives states, to elicit particular reactions.[2]

Brown's restriction of the use of the terms "motive" and "motivational" to refer to dynamogenic functions alone accords with Hull's recent formulations, as discussed by Spence (**47**, **48**), and is believed to be in the best interests of terminological clarity. However, these terms will be used in the present paper in their more traditional (and ambiguous) meaning, to refer to states or processes having both directing and dynamogenic components. The directive aspects of motives will be referred to variously as "associative," "cognitive," or "habitual," whereas the term "drive" will be used to refer to the nondirective, energizing and reinforcing aspects, in accordance with Hull's usage of the concept *D*.

Since certain variables not usually classified as motivational, e.g., ordinary environmental stimuli, may have strong associative connections even though they have little or no drive value, it is assumed that it is their drive function, basically, that distinguishes motivational variables from other kinds. Thus, the mere demonstration that a given variable has directive characteristics is not enough to qualify it as motivational. Furthermore, even if a variable qualifies as motivational because of its evident energizing or reinforcing effects under some circumstances, its effects upon behavior in other circumstances may depend primarily upon its associative properties, i.e., its tendency to elicit particular responses. On the other hand, the fact that a variable has associative properties does not preclude the possibility of its having drive properties as well.

The implications of this general view are here considered in respect to the interpretation of the role of certain variables, commonly regarded as motivational, in verbal learning and performance. Although the discussion is somewhat arbitrarily limited to only a few such variables in relation to human verbal behavior, it is believed that these variables are not unrepresentative of motivational variables in general, and also, that the considerations raised concerning the analysis of their effects apply as well to nonverbal behavior, in both animals and men (cf. **15**).

---

[2] Postman has suggested that the notion of a nondirective drive may make the concept of drive expendable. Thus, he proposes that the explanation of variability in behavior, despite the constancy of stimulus situation and habit strength, might well be sought in the "changing patterns and intensities of drive stimuli" (**37**, p. 57). This sort of formulation, in point of fact, bears a close resemblance to that of Dollard and Miller (**13**), who define a drive as a strong stimulus. It does not, however, as Postman further suggests, endanger the Hullian distinction between $_{S}H_{R}$ and $_{S}E_{R}$, nor for that matter, the somewhat similar Tolmanian distinction between cognition and performance. Hull's D (as opposed to S_D) *is* nondirective; a concept like drive would presumably still be needed to explain at least some kinds of reinforcement phenomena; and intensity of stimulation, in Hull's last formulations, is itself considered a motivational variable multiplying habit strength (cf. **47**, **48**).

Experimentally Manipulated (Environmentally Defined) Motives

Electric shock. In a large number of investigations of the effects of motivation, the state of *S* is more or less under the control of the investigator. These "experimentally manipulated" states have, for the most part, been acquired. However, one primary motivational state whose effect upon verbal learning and performance has received considerable attention is that induced by electric shock.

To say that the effects of electric shock are not very well understood is an understatement. Nevertheless, an examination of studies involving its use is instructive, since it indicates many of the considerations that must attend analyses of so-called motivational variables in general. First of all, it seems clear that electric shocks qualify as motivational by virtue of their drive-producing effects. The role of electric shock as a reinforcer is well known, and its general energizing effect has been demonstrated in a number of studies (e.g., **1**, **35**, **42**, **51**). But these are by no means its only functions. For instance, electric shocks may be informative. Thus, under certain conditions, particularly when appropriate verbal instructions are given, electric shock may lead to improved performance, whether administered for wrong responses or for right ones (**28**). It is probably gratuitous, and even misleading, in these instances, to attribute the effects of the shock solely to its drive characteristics. Since shock has not only drive value, but also cue value, its associative or cognitive aspects must be taken into account.

Even in those instances in which shock is noninformative, i.e., not related in any systematic way to correct or incorrect responses, it is necessary to consider its response-producing properties. The fact that shock tends to elicit withdrawal, or vocalization, or muscular contraction in rats or human beings, may be a consequence, not of its drive characteristics, but rather, of its association with either learned or innate response tendencies. Since shock is an intense stimulus, capable of eliciting a variety of responses, it is not surprising that in certain situations it may elicit responses that are incompatible with those elicited by other stimulus components of the situation. Certain kinds of verbal performance are perhaps particularly susceptible to disruption (interference) by shock, though there seem to have been no investigations specifically designed to discover whether verbal behavior is more susceptible in this respect than nonverbal behavior of various kinds, particularly when other factors, such as level of practice, are held constant.

Although the disrupting effect of shock, when it occurs, is often attributed to the disorganizing influence of emotion, it seems unnecessary, at least in some situations, to attribute the interference to anything more than the introduction of an intense, extraneous stimulus which produces strong competing responses. This interpretation may also apply to other kinds of prepotent stimuli that have a distracting or disrupting effect, e.g., loud noises, loss of support, or simply talking to *S*.

It must be noted that some kinds of verbal behavior, e.g., swearing, are likely to be elicited by shock or other noxious stimuli as a result of learning. If such responses are considered "correct" in verbal learning situations, shock might be expected to facilitate performance. Depending, then, upon the particular responses that have been acquired and the particular requirements of the test situation, shock might be either beneficial or detrimental. Unfortunately, little is known concerning the relative strength of various kinds of verbal response to shock. It does not seem improbable, however, that at least some of the effects of shock are attributable to learned associations rather than to its drive value as such.

Hunger. A somewhat more confident assessment of the role of stimulus cues may be made in relation to certain studies of the effects of the primary drive of hunger. Thus, there is evidence, not without its exceptions, to be sure, that increasing hunger increases the probability of occurrence of verbal responses referring to food. When this does occur, as both Brown (**5**) and Postman (**36**) have pointed out, it is most easily explained on the supposition that responses relating to food are more strongly associated with hunger stimuli than are most other responses.

This is not to say there is no evidence that hunger, defined in terms of food deprivation, functions as a drive. Studies showing that the receipt of food following deprivation is reinforcing, or that the presence of hunger has a generally enhancing effect upon response tendencies, attest to its drive properties. But when the responses elicited or intensified in the presence of a drive are "need-related," i.e., when they are responses that have been conditioned to the cues characteristic of a given drive, their explanation obviously requires no appeal to drive *qua* drive. Studies showing that the frequency of verbal responses relating to food or eating in the presence of impoverished or ambiguous verbal or visual stimuli increases as a function of hunger are interesting and important. But they appear to deal, for the most part, with the associative rather than the drive properties of hunger.

Fear (anxiety) resulting from noxious stimulation. Many verbal learning studies involving electric shock are designed to investigate its effects upon subsequent performance under nonshock conditions. It is not improbable that behavior under these conditions is in part a function of purely perseverative effects of pain, or what a number of investigators (e.g., **1**, **42**, **43**) have termed "emotionality." However, the motivational state usually considered in this connection is that presumably aroused by environmental cues associated with the shock. The state thus aroused is an acquired one, namely, fear or anxiety. The demonstration that fear, so defined, has nonassociative properties rests primarily, thus far, upon animal experiments indicating that the reduction of fear is reinforcing (**31**), or that its occurrence intensifies ongoing responses that are not themselves produced by the fear (**7**).

As in the case of hunger, many of the experimental studies of the effects of fear on human behavior involve the learning or retention of motive-related materials, i.e., particular words or other items that had previously appeared in close temporal contiguity with a noxious stimulus. As a result of conditioning, such items might be expected to elicit not only fear, but also some of the other responses previously elicited by the shock. But in addition, as Dollard and Miller (**13**) point out, fear itself brings with it a number of response tendencies, either learned or unlearned. (This holds true, of course, even though the fear is produced by stimuli other than those comprising the material to be learned or recalled.)

The empirical evidence concerning the effects of fear on verbal performance has not been altogether consistent. Under certain conditions the production of fear seems to facilitate the recall or recognition of verbal materials (**12, 34**). More often, however, it seems to elicit responses that successfully compete with verbal behavior. Considered in this manner, fear-produced responses might well be classed among those more prosaic variables usually considered as determinants of associative interference or facilitation. We have, as yet, very little information concerning the kinds of response tendencies ordinarily produced by fear, or the way in which they interact with other response tendencies in specific situations. There is reason to believe that fear may have particularly intense stimulus properties, and, other things being equal, may therefore elicit unusually strong responses, of a sort not ordinarily observed in laboratory studies of forgetting (cf. **40**). But this does not in itself necessitate an explanation of its effects in other than associative terms. Again, it must be emphasized that this is not to deny that fear acts as a drive. It does deny that disruption (or for that matter, facilitation) of behavior by fear is in itself an adequate basis for the supposition that it has drive properties (**15**).

Anxiety aroused by socially tabooed materials. Somewhat related to investigations using items associated with noxious stimulation are the studies involving socially tabooed verbal materials. In these studies it is supposed that certain words or phrases arouse an acquired anxiety drive, not because of their association with punishment in the experimental situation, but because of previously experienced social punishments, administered either for the overt usage of the verbal expression itself, for behavior labeled by the verbal expression, or for behavior associated with stimuli thus labeled. Anxiety aroused in this manner is sometimes regarded as more similar to the sort of anxiety considered by clinicians as a determinant of repression than that produced by electric shock (**40**).

In many of the studies in which it has been found that the learning, retention, or recognition of socially tabooed or negatively valued items is impaired, the interpretation is often in terms of unconsciously motivated defense mechanisms. This sort of interpretation has been seriously challenged on two grounds: first, that the mechanism involved is typically far from unconscious; and secondly, that it primarily involves habitual rather than purely motivational mechanisms.

The first objection stresses the possibility that the overt expression of socially unacceptable responses is at least sometimes deliberately and consciously inhibited (e.g., **21, 36, 57**). The second objection rests on the finding by Howes and Solomon (**22**) that the visual recognition thresholds for words vary with their relative frequency of usage, as determined by the Thorndike-Lorge word counts. By taking account of the large difference in the mean frequencies of usage of the taboo and neutral words used in an investigation by McGinnies (**29**), Howes and Solomon (**21**) were easily able to account for McGinnies' finding that the taboo words had the higher thresholds. Postman also has adopted the view that high thresholds for negatively valued words are explicable in terms of their relative habit strengths. Pointing out that the report of such words may be inhibited by the tendency to give more frequently used neutral words that have elements in common with the taboo words, he has concluded that

> . . . it is not necessary to ascribe selective functions to the immediate motivational conditions of the perceiver. Defensive and/or vigilant thresholds as well as systematic prerecognition hypotheses may be consequences of the relative strengths of response dispositions aroused by specific stimuli (**36**, p. 80).

This interpretation, it should be noted, is not inconsistent with the view that socially disapproved words may arouse anxiety, and that the anxiety acts as a drive. It is not improbable that the low frequency of usage of these words is itself due to their evocation of anxiety and to the reinforcement of their suppression resulting from anxiety reduction.[3]

Failure defined by verbal instructions. One of the most widely used techniques for inducing so-called motivational changes in verbal learning situations is the use of instructions by *E* indicating that *S* has failed. As Lazarus, Deese, and Osler (**26**) have observed, a distinction must be made between the purely motivational (drive ?) effects of failure instructions and their informative or distracting effects. Thus, the problems of distinguishing between the drive and the associative components of motivational variables must be considered in relation to failure experiences also.

Experimental studies have usually tended to show that when failure instructions are not applied to specific responses, but are used, instead, to characterize general performance, subsequent performance in verbal learning situations is impaired. As a result of such observations, it is generally believed that statements expressing approval of a learner's performance are more beneficial to his subsequent performance than statements indicating disapproval. At the same time, however, it is also rather generally believed that excessive praise may decrease motivation and thus impair performance, whereas some kinds or degrees of censure may increase motivation and thus improve performance. This indecision with respect to the effects of failure characterizes treatments of the effects of frustrating events in general. For instance, Child and Waterhouse write:

> Indeed, the conflict appears strikingly in some general textbooks in psychology. In a chapter on thinking and reasoning frustration is viewed as the condition for more organized behavior, and in a chapter on emotion it is viewed as the condition for less organized behavior. The failure to use a common term such as frustration in the two chapters apparently permits the contradiction to go unnoticed (**10**, p. 127).

In their subsequent discussion, these writers have presented a notably useful analysis of the manner in which the responses produced either directly or indirectly by the frustrating event might either improve or impair performance. They suppose that prevention of the dominant response in the hierarchy may lead to its extinction, with the result that initially weaker responses are now evoked; or the frustrating event may modify the situation in such a way that some new response is elicited, either by some changed element of the situation or by the frustrating event itself. These possible effects of frustrating events, it should be noted, relate exclusively to the modification of cues and/or to the relative strengths of habits associated with these cues.

---

[3] This account is clearly neutral in respect to the question of the usefulness of the concept of "perceptual defense." It affirms the possibility that anxiety cues may produce responses that successfully compete with correct recognition, and is altogether in accord with the view that there may be marked individual differences in the amount of anxiety aroused by various cues and/or the nature of the responses elicited by this anxiety. Furthermore, following Dollard and Miller (**13**), it freely admits the possibility that the individuals concerned may sometimes be unconscious of the cues eliciting the anxiety.

This emphasis by Child and Waterhouse upon the associative aspects of frustration extends also to their treatment of the possible drive-producing consequences of frustrating events. Thus, they suppose that ". . . frustration may operate to increase the motivation supporting the goal-oriented activity and thereby to improve the quality of performance" (**10**, p. 136), or conversely, to produce a drive state of intense general emotion which impairs performance either ". . . because the emotional responses themselves are to some extent incompatible with the ongoing instrumental activity" (**10**, p. 137), or because the emotion produces other incompatible responses. It would appear, then, that the only properties explicitly attributed by these writers to the motivational states that might be produced by frustrating events are directive.

They suggest, correctly, that the foregoing treatment has something in common with the "emotional" interpretation of frustration behavior by Brown and Farber (**6**). But it seems, in general, to fall more clearly within the class of "nonemotional" interpretations, as discussed by the latter writers, since it deals with the possible behavioral consequences of frustration without appealing to the concept of a frustration-produced drive having nonassociative properties. As a matter of fact, until recently there has not been much evidence of a positive nature to support the widespread notion that frustration produces (or acts as) a drive. Consequently, it has been doubtful whether, from the present point of view, frustrating events qualify as motivational variables (cf. **6**). However, recent studies by Amsel and Roussel (**2**) and by Lambert and Solomon (**25**) indicate that at least some kinds of frustrating events affect animal behavior in the manner characteristic of a drive.

Whether failure, defined in terms of verbal instructions, has a comparable effect upon verbal behavior still remains to be demonstrated. Unfortunately, under most conditions of verbal learning, there is no easy way to separate associative from possible drive effects of supposedly motivational variables. Contrary to general belief, the impairment of performance by failure does not necessarily demonstrate its associative effects, nor does the improvement of performance by failure demonstrate its drive effects. Just as the associative mechanisms attending failure may either benefit or harm performance, so might an increase in drive, in and of itself, affect behavior either favorably or adversely, depending on the specific nature of the task and the experimental conditions involved (**16**, **48**). Since few, if any, studies of the effects of failure have thus far been designed in a manner that permits an evaluation of its drive properties, the motivational status of this particular variable still seems somewhat doubtful.[4]

---

[4] Evidence concerning the reinforcing effects of escape from or avoidance of failure must not be neglected, of course, in considering this

Differentially Selected (Response-Defined) Motives

Logic of response-defined hypothetical constructs. The variables discussed so far have consisted of directly manipulable environmental conditions. Very often, however, motivational states are defined in terms of performance measures of one sort or another. In order to observe the effects of variations in motivation, thus defined, one must differentiate or select individuals who vary in terms of the particular response index chosen. If the response measure does reflect degree of drive, and if one has a theory concerning the way in which such variations influence behavior, information concerning this response index for individual *S*s should enable one, in specific situations, to predict their behavior more accurately than one could do otherwise.

As Spence has pointed out (**45**), the usefulness of response-inferred theoretical constructs is in some ways limited, and, in principle at least, it should be possible, given sufficient information concerning antecedent conditions, both historical and contemporaneous, to explain or predict the very behavior used as the index of the hypothetical state under consideration. However, when information concerning these antecedent conditions is meager or unavailable, the use of responses as indices of intervening variables is clearly advantageous. The importance of response-inferred constructs has been particularly stressed by those psychologists who insist that behavior cannot be predicted at all without ascertaining the "meaning," or *S*'s "perception" of a situation. In practice, they use certain responses, usually verbal, as indices of a "psychological environment," or "phenomenological field," or some other hypothetical structure, to which they assign a primary explanatory role. From an S-R point of view there can be no quarrel with this procedure, provided that the responses from which the hypothetical states are inferred are explicitly and unambiguously identified. However, as Bergmann observes, it is " . . . the business of science to ascertain which objective factors in the past and present states of the organism and its environment account for the difference in response . . . " (**4**, p. 133), including, presumably, the response(s) from which one infers the "meaning" for *S* of any given situation.

Of course, statements of relations between one kind of behavior and another, which Spence (**45**, **46**) calls "R-R laws," need not involve any theory about motivation, or, for that matter, any other hypothetical state. But if a given response measure is labeled an index of motivation, and if this labeling is to serve any useful purpose, one's theory concerning the relation of motivation to behavior ought to be made explicit. Otherwise, there is no justification for considering the response in question to be a measure of motivation rather than anything else.

It is instructive, from this point of view, to examine the many response measures that have been proposed as indices of motivational states. When this is done, one finds a very meager basis indeed for supposing that the majority reflect anything but associative tendencies. Although such findings do not in any way minimize the importance of these instruments or the empirical relations established in connection with them, they empha-

---

problem. It seems probable that failure or anticipation of failure can be profitably conceptualized in terms of a state resembling anxiety, which, when defined in certain ways, seems clearly to qualify as a motive (**15**).

size the necessity for formulating adequate criteria of motivation.

Allport-Vernon Study of Values. The problem of the motivational status of response-defined variables has received a good deal of attention recently in the specific case of performance on the Allport-Vernon Study of Values. As a result of the able analyses by Solomon and Howes (**44**) and Postman (**36**), it seems evident that variations in scores on this particular instrument are much more reasonably interpreted in terms of variations in the habit strength of certain verbal responses than in terms of variations in drive factors per se. Studies by these investigators have demonstrated empirically that the relative frequency of usage of words in the various value areas comprising the Allport-Vernon scale is the most important determinant of recognition thresholds for words associated with these value areas.[5] But on theoretical grounds alone it could be concluded that a response index that does nothing but predict the relative strength of other specific responses, and particularly those highly related to the items comprising the index, does not qualify as a measure of drive. There is no question that the Allport-Vernon Test measures something; but, from the present point of view, there is no reason to suppose that the "something" is a motive.

Achievement imagery. Recently, McClelland, Atkinson, Clark, and Lowell (**27**) have published an important analysis of motivation, centered about their experimental investigations and those of their colleagues and students, of achievement imagery, as measured by the contents of stories given in connection with TAT-like pictures. The results of these investigations have indicated that *S*s who vary in their achievement imagery scores tend to differ, under certain conditions, in various other kinds of performance, including verbal. The predictive usefulness of this behavior index is clearly attested to by such findings, but it is not entirely certain that they demonstrate that the test reflects motivational characteristics.

Thus, the obtained relations between achievement imagery scores on the one hand, and visual recognition thresholds for achievement-related words or the frequency of usage in a written essay of certain terms (see **27**, p. 252) on the other, seem to be precisely those that would be expected on the assumption that variations in achievement scores reflect variations in the strengths of certain verbal *habits*. In other words, the interpretation by Solomon and Howes and by Postman of the behavioral correlates of scores on the Allport-Vernon Scale may apply with equal cogency to these results also.

The finding that increment of output with practice in a Scrambled Words Task was positively related to achievement scores is given particular weight by McClelland and his colleagues in arguing that scores on their test are measures of motivation. They point out that, for some theorists, "... the decisive criterion for determining whether a motive is involved in performance is whether or not it can produce learning. Since there is clear evidence of learning in the high *n* Achievement group, it can be argued that the *n* Achievement score is a measure of motivation" (**27**, p. 234). From the present point of view this reasoning is not very convincing. Although motivation is held to be essential to learning, it does not follow from this assumption alone that, with other factors presumably held constant, better performance in learning situations is necessarily due to a higher drive level. Indeed, Hull's theory leads to the prediction that, under certain learning conditions, *S*s with higher drive will perform more poorly than those with lower drive (**16**, **33**, **38**, **48**, **54**).

---

[5] This does not mean that variations in the habit strength of the verbal responses might not themselves be attributable to variations in drive. This possibility, however, would require independent verification.

The demonstration that the amount of food-related content in *S*s' stories tends to be positively related to number of hours of food deprivation has also been used by the Wesleyan group as evidence of the influence of drive factors upon the kind of imaginative productions from which they have derived their *n* Achievement scores. Since hunger, defined in terms of number of hours of food deprivation, qualifies as a drive by virtue of the demonstration, in other studies, of its enhancing and reinforcing properties, it is possible that the amount of food-related content in TAT stories reflects the drive aspect of hunger directly. But, as the discussion in an earlier section of this paper has indicated, an alternative interpretation in terms of the verbal habits associated with hunger is also clearly possible.

In general, the authors of *The Achievement Motive* assume that responses in fantasy situations are relatively unaffected by variations in habitual or cognitive factors:

> That is, for most subjects putting thoughts into words or verbalizing is a highly overlearned response [and is therefore a "normal" habit, whose strength is substantially similar in most individuals]. Furthermore, in the fantasy situation no particular set of responses is supposed to be perceived as especially appropriate. Fantasy is a "free" response situation, provided the picture is not too structured. It might not be for a certain class of persons, for professional writers, for example, because they may have learned a particular set of responses to use in such a situation. . . . But except for professional authors, individuals should have no particular set of verbal response tendencies which seem appropriate because of past experience with such situations (**27**, pp. 40–41).

Although it is somewhat difficult to evaluate the merits of this argument, since the meaning of the concept, "normal habit," is not altogether clear, there seems to be plenty of evidence that stimulus factors and specific past experiences play at least some role in determining TAT performance (e.g., **11**, **14**, **39**). The factors determining the kind and amount of influence of associative variables upon performance in fantasy situations like that presented by the TAT are certainly not so well known as those determining the kinds of verbal performance traditionally studied in the laboratory. At present, therefore, the supposition that variations in performance on so-called "projective" tests do not reflect variations in associative factors, like the assumption that they do, appears to owe a greater debt to faith than to empirical fact.

Iowa Picture Interpretation Test (IPIT) achievement scores. Despite the foregoing demurrals, the extensiveness of the range of responses apparently affected by the achievement variable, as defined by the Wesleyan group, lends some credibility to the supposition that achievement imagery scores may indeed reflect strength of drive. Among the studies of verbal behavior, the most important, in this connection, are those that demonstrate a relation between achievement scores and other kinds of verbal behavior whose content has no obvious relation to that from which the achievement scores are themselves derived.

If achievement imagery scores are positively related to strength of drive, it could be predicted, on the basis of Hull's (**23**) assumptions concerning drive combination, that the introduction of an experimental variable calculated to increase drive would have less effect on the behavior of *S*s with high achievement scores than it would have, in general, on the behavior of those with low achievement scores (cf. **6** for a similar deduction in regard to frustration). Stated somewhat differently, a correlation might be expected under ordinary conditions between achievement imagery scores and performance. The addition of another drive variable, however, ought to reduce this correlation.

To test this hypothesis, Hurley (**24**) recently administered the IPIT to a large number of *S*s. In this test, *S*s are given a choice of four interpretations, or imaginative descriptions, of each of ten TAT pictures. One of the four interpretations of each picture involves achievement orientation, as judged by clinical psychologists on the basis of a definition taken from *The Achievement Motive*.[6] The other three interpretations for each picture were designed, respectively, to indicate hostility, insecurity, and blandness.

When *S*s were given instructions intended to be poorly motivating, a significant positive relation was found between achievement scores and the number of correct anticipations in a serial rote learning task. Under high motivational instructions, however, the correlation was insignificant.[7] There was also some tendency for achievement scores to be positively correlated with number of erroneous anticipations under the low motivational instructions. This result, though not statistically significant, was similar to a finding reported by Lowell (see **27**).

Although Hurley's results were in accord with predictions made on the assumption that achievement scores reflect degree of motivation, they are also consistent with the notion that achievement scores reflect a general response tendency influencing performance in a wide variety of situations. *S*s with high achievement scores may simply have the habit of being or at least appearing "eager." In other words, they may have learned certain work habits of an "effortful" or "active" sort. The distinction between drive-produced activity and learned activity has been nicely brought out by Miller (**32**), who points out that rats can be *trained* to be extremely active when hungry if pellets of food are widely scattered about a maze. Animals trained in this way would presumably be more active ("eager-appearing") than untrained animals, even though the latter might actually be hungrier. In like manner, human *S*s appear to vary, as a result of learning, in the degree of obviousness of their desire to succeed. Under low motivational instructions *S*s having relatively strong habits of this sort should tend to produce both more correct and incorrect responses. When the instructions themselves produce the "eagerness" response, the differences between the two groups might be expected to disappear. It is far from certain, then, that the term "achievement motive" is appropriate, at least when applied to the indices here considered; "achievement-oriented habit," though more awkward, may be more accurate.

IPIT hostility scores. Certain of

---

[6] This was made possible by the generous cooperation of Dr. J. W. Atkinson, who furnished a prepublication copy of the manuscript.

[7] The difference between the two correlations was significant at $p<.07$. The higher level of significance reported for this comparison in the original version of this paper was based upon a smaller N.

Hurley's results suggest that the personality (response-defined) variable operating in his study may actually have been hostility, rather than achievement-orientation. He found a highly significant negative correlation between the hostility scores on the IPIT and the number of both correct and erroneous anticipations in verbal learning, under low motivational instructions. This correlation disappeared under the highly motivating instructions, the difference between the two correlations being significant. Since the correlation between scores on the achievement and hostility indices of the IPIT was high ($r = -.71$), the relation of each score to performance was evaluated by using partial correlations. When the effect of the hostility score was eliminated, the partial correlations between achievement and performance were reduced to insignificance. However, when the effect of achievement imagery scores was partialled out, the negative correlation between hostility and number of correct and total (correct plus incorrect) anticipations under the low motivational conditions remained significant.

It is possible that high hostility, as measured by the IPIT, itself reflects a drive state. Since high *D* may, under certain conditions, be expected to produce poorer verbal performance (see below), such an assumption would not be inconsistent with Hurley's results. But this would imply that high achievement scores reflect low *D*. It seems simpler and more plausible, as Hurley himself concluded, to attribute the relations between the hostility scores and performance to differences in habit tendencies rather than to drive.

Taylor Anxiety Scale (A-*scale*) *scores*. A number of investigators in the Iowa Laboratory and elsewhere have been expressly concerned with the experimental and theoretical problems involved in testing the supposition that a given response measure reflects drive. Using an anxiety questionnaire originally constructed and recently revised by Taylor (**52**, **53**), these investigators reasoned as follows: If *S*s' scores on this test reflect general drive level, and if variations in drive affect response strength in the manner suggested by Hull (**23**), then the performance of these *S*s should differ in certain predictable ways in given experimental situations.

Since the studies were for the most part deliberately restricted to a consideration of nonassociative characteristics on the view that this procedure is most appropriate in a test for motivational properties, the assumption figuring most prominently in these predictions was that all habit tendencies elicited in a given situation are multiplied by the value of the total effective drive then operating. On this assumption, it can be shown that in learning situations in which correct responses are highly dominant, *S*s with a relatively high drive level should perform better than those having a low drive level.[8]

If, however, erroneous response tendencies are stronger than the correct ones, the performance of *S*s with high drive should be poorer than that of *S*s with low drive. Furthermore, as Taylor and Spence (**54**) have pointed out, in complex situations, even though a correct response tendency might be slightly stronger than any alternative superthreshold tendencies, an increase in drive could nevertheless increase the frequency of erroneous responses. This does not mean that the presence of competing tendencies would always put *S*s with high drive at a relative disadvantage, regardless of the strength of the correct tendencies, but, rather, that the relative strengths of the correct and incorrect tendencies are not the only factors to be considered. Thus, according to Hull's theory, an increase in drive (*D*) will increase the excitatory strength (*E*) of the strongest response more than that of the weaker ones, but will also increase the probability that more of the weaker responses will have superthreshold values. If the increase in *D* does not increase the *E* of the strongest tendency so much that its downward oscillation never brings it within the range of *E*'s for the weaker responses, the possibility of successful competition by weaker, but now superthreshold tendencies, may be increased. Clearly, there is no possibility of successful competition by the weaker tendencies if their strengths are subthreshold.

---

[8] Contrary to the supposition by Lazarus, Deese, and Osler (**26**), Taylor's prediction that anxious *S*s would give more conditioned eyelid responses than nonanxious *S*s was based upon this multiplicative assumption ($R = H \times D$), rather than upon the assumption that anxious *S*s have the stronger associative tendency to make defensive reactions. In the case of simple eyelid conditioning, it is not possible, of course, to decide between these two interpretations of Taylor's original results (**52**). In other situations, however, the behavior predicted on the basis of these different hypotheses would presumably be quite different.

A number of deductions that could be made on the basis of the foregoing theoretical analysis have been tested in various situations, including verbal learning (**33**, **38**, **48**). The results have indicated that the quality of performance in complex learning situations is inversely related to *S*s' degree of anxiety, as measured by the Taylor scale, and furthermore, that the advantage of the nonanxious over the anxious *S*s is positively related to the probable number and strength of the competing responses elicited. At the moment of this writing it does not seem altogether certain whether conventional verbal learning tasks can be made simple enough to provide a consistent advantage for the anxious *S*s. For instance, Montague's (**33**) anxious *S*s were superior to the nonanxious on an easy list, consisting of items having high association values and a minimum of intralist similarity, but this simple effect was not significant; only the interaction between anxiety and difficulty of list was significant. However, more recent studies (see **48**) indicate that it is possible to construct verbal learning situations that permit highly anxious *S*s to reap the benefit to which their higher drive, if it indeed exists, should entitle them under appropriate conditions.

Theoretical implications of A*-scale studies.* It has not always been recognized that most of the Iowa studies of the relations between *A*-scale scores and various kinds of laboratory performance have been exclusively concerned with the test of the possibility that the *A* scale measures drive level (*D*). From the present point of view, the test of the supposition that the *A* scale or any other variable is a measure of *D*, or of motivation, generally, must be made in terms of predictions based on purely nonassociative assumptions, a procedure seldom if ever followed in the case of other response indices. It must be noted that this kind of test can be made only under certain conditions, namely, those in which the associative properties of the variable may be presumed to have little differential effect. Under such necessarily limited conditions, the general success of the predictions thus far may fairly be said to lend increased confidence in a drive interpretation of *A*-scale scores.

This conclusion does not imply that *A*-scale scores reflect nothing but *D*. On the contrary, if they reflect a motivational state, then in the present view, they also undoubtedly reflect associative tendencies. Under many circumstances, therefore, *S*s with relatively high scores should perform differently from those with low scores by virtue of differences in their habits. In such instances one could not expect to account for variations in the behavior of anxious and nonanxious individuals in terms of drive level alone.[9]

The theoretical studies concerned with the *A* scale were never intended, then, to investigate all of its possible functional relations. Nor, contrary to the belief of some, was the *A* scale constructed in order to predict as well as possible either eyelid conditioning, or verbal learning, or motor skill, or any other sort of performance. If this had been so, a very different procedure would undoubtedly have been followed. If one wished to use a self-rating questionnaire, one would administer a very large number of items, obtain the correlation of each with the behavior to be predicted, and then, after cross validation and a consideration of the item intercorrelations, select the set that most efficiently predicted the criterial measure. A test constructed in this manner would unquestionably predict the specific kinds of behavior it was designed to measure better than the *A* scale could be expected to do.

It is certainly not surprising that various kinds of empirically selected measures excel the *A* scale as predictors of certain kinds of behavior. But since the *A* scale was not devised, nor ever considered by those associated with its construction, to be a measure of all of the variables influencing performance, some sort of limit is to be expected, under most conditions, in the amount of variance in performance that is associated with scores on the *A* scale. It is only in certain kinds of controlled situations that variations in performance are likely to be primarily due to variations in drive.

A misconception somewhat related to the notion that the *A* scale was intended to be a "best predictor" of some kind is that the "essential" validity of the Taylor *A* scale depends upon its correlation with other indices of anxiety, e.g., those of a clinical or psychiatric nature. There is some evidence, as a matter of fact, that scores on this test do relate to clinically diagnostic measures, including psychiatric ratings (**17**). Such relations, when they are found, are interesting and useful (though perhaps not so useful as one might wish, in view of the notorious unreliability of psychiatric judgments), but they most emphatically do not constitute a validation of the test in respect to the purpose it was originally designed to serve. As Bechtoldt (**3**) has recently observed, the question of the validity of the Taylor scale *as a useful definition of general drive level* is answered by the accuracy of the prediction of relations between this scale and specified behavior variables, under conditions such that variation in the behavioral variables can be reasonably attributed to differential drive levels.

---

[9] From the standpoint of these considerations, Child's (**9**) criticisms of the Iowa studies for failing to investigate the associative aspects of manifest anxiety do not seem to have been well-taken. Although the motivational status of anxiety could not be demonstrated by pointing to differences in the habits of anxious and nonanxious *S*s, there is no doubt that such differences exist. On the other hand, despite considerable speculation and some empirical evidence (e.g., **3**, **18**, **19**, **55**, **56**), it is not as yet entirely certain what these habitual differences may be, apart from the trivial observation that they consist, at least in part, in the kinds of verbal responses given on the test itself. For instance, despite some supporting evidence (**20**), it is doubtful, on either theoretical or empirical grounds, (**49**, **50**) that a consideration of the specific nature of anxiety responses warrants the belief that anxious *S*s are (habitually ?) poorer than nonanxious *S*s in discriminating between threatening and nonthreatening stimuli.

In view of this, the Taylor questionnaire might better have been termed a test of "excitability," or "level of responsiveness," or simply, in accordance with the present usage, an "*A* scale." Nevertheless, the items of the scale were selected by clinical psychologists in accordance with the specifications for manifest anxiety laid down by a descriptive psychiatrist of high repute (**8**). It does not seem inappropriate, consequently, to consider this test an index of manifest anxiety as well as *D*.

SUMMARY

It is the major thesis of the present paper that the question as to whether a given variable acts as a motive or reflects a motivational state can be answered only in the context of explicit assumptions concerning the particular ways in which motivational variables can be differentiated from other variables in their influence on behavior. In view of the current confusions between the concepts of motivation and habit, it would appear to be particularly desirable to distinguish between the associative and nonassociative properties of variables. Thus, experimentally manipulated or environmentally defined variables such as electric shock, hunger, fear, and failure may be classed as motivational by virtue of their nonassociative (drive) properties; nevertheless, their effects upon verbal learning and performance in some situations are best understood in terms of their associative characteristics. With a very few exceptions, the influences of differentially selected or response-defined variables frequently classified as motivational also seem attributable, for the most part, to their associative properties; in the absence of convincing evidence that such variables have anything other than an associative influence, it is doubtful whether they qualify as motivational at all.

REFERENCES

1. AMSEL, A., & MALTZMAN, I. The effect upon generalized drive strength of emotionality as inferred from the level of consummatory response. *J. exp. Psychol.*, 1950, **40**, 563–569.
2. AMSEL, A., & ROUSSEL, JACQUELINE. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.
3. BECHTOLDT, H. P. Response defined anxiety and MMPI variables. *Proc. Ia. Acad. Sci.*, 1953, **60**, 495–499.
4. BERGMANN, G. Psychoanalysis and experimental psychology: a review from the standpoint of scientific empiricism. *Mind*, 1943, **52**, 122–140.
5. BROWN, J. S. Problems presented by the concept of acquired drives. In *Current theory and research in motivation.* Lincoln, Nebr.: Nebraska Univer. Press, 1953. Pp. 1–21.
6. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–480.
7. BROWN, J. S., KALISH, H. I., & FARBER, I. E. Conditioned fear as revealed by magnitude of startle to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317–328.
8. CAMERON, N. *The psychology of behavior disorders.* New York: Houghton Mifflin, 1947.
9. CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–171.
10. CHILD, I. L., & WATERHOUSE, I. K. Frustration and the quality of performance: II. A theoretical statement. *Psychol. Rev.*, 1953, **60**, 127–139.
11. COMBS, A. W. The use of personal experience in Thematic Apperception Test story plots. *J. clin. Psychol.*, 1946, **2**, 357–368.

12. Diven, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.
13. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
14. Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, **12**, 387–395.
15. Farber, I. E. Anxiety as a drive state. In *The Nebraska symposium on motivation.* Lincoln, Nebr.: Nebraska Univer. Press, 1954.
16. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
17. Gleser, Goldine, & Ulett, G. The Saslow Screening Test as a measure of anxiety-proneness, *J. clin. Psychol.*, 1952. **8**, 279–282.
18. Goodstein, L. D. Interrelationships among several measures of anxiety and hostility. *J. consult. Psychol.*, 1954, **18**, 35–39.
19. Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–454.
20. Hilgard, E. R. Theories of human learning and problems of training. In *Symposium on psychology of learning basic to military training problems.* Panel on Training and Training Devices. Res. & Dev. Bd., 1953.
21. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
22. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
23. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
24. Hurley, J. R. Verbal learning as a function of instructions and achievement motivation. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
25. Lambert, W. W., & Solomon, R. L. Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, 1952, **45**, 269–279.
26. Lazarus, R. S., Deese, J., & Osler, Sonia F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
27. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
28. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (2nd Ed.) New York: Longmans, Green, 1952.
29. McGinnies, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1946, **56**, 244–251.
30. Melton, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* (2nd Ed.) New York: Macmillan, 1950. Pp. 668–690.
31. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
32. Miller, N. E. The role of motivation in learning. In *Symposium on psychology of learning basic to military training problems.* Panel on Training and Training Devices. Res. & Dev. Bd., 1953, 103–116.
33. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
34. Murphy, D. B. Recognition thresholds and recall of nonsense syllables as a function of their frequency of occurrence, their after-effects, and of personality variables. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
35. Nagaty, M. O. The effect of reinforcement on closely following S-R connections: I. The effect of a backward conditioning procedure on the extinction of conditioned avoidance. *J. exp. Psychol.*, 1951, **42**, 239–246.
36. Postman, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation*. Lincoln, Nebr.: Nebraska Univer. Press, 1953, pp. 59–108.
37. Postman, L. Comments on papers by Professors Brown and Harlow. In *Current theory and research in motivation*. Lincoln, Nebr.: Nebraska Univer. Press, 1953. Pp. 55–58.
38. Ramond, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
39. Roe, Anne. A psychologist examines 64 scientists. *Sci. Amer.*, 1952, **187**, 21–25.
40. Sears, R. R. Functional abnormalities of memory with special reference to amnesia. *Psychol. Bull.*, 1936, **33**, 229–274.
41. Shaffer, L. F. *The psychology of adjust-*

ment. New York: Houghton Mifflin, 1936.

42. SIEGEL, P. S., & BRANTLEY, J. J. The relationship of emotionality to the consummatory response of eating. *J. exp. Psychol.*, 1951, **42**, 304–306.
43. SIEGEL, P. S., & STUCKEY, HELEN S. The effect of emotionality on the water intake of the rat. *J. comp. physiol. Psychol.*, 1949, **42**, 12–16.
44. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thesholds. *Psychol. Rev.*, 1951, **58**, 256–270.
45. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.
46. SPENCE, K. W. The postulates and methods of 'Behaviorism.' *Psychol. Rev.*, 1948, **55**, 67–78.
47. SPENCE, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.) *Comparative psychology.* (3rd Ed.) New York: Prentice Hall, 1951.
48. SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Learning theory, personality theory, and clinical research: the Kentucky symposium.* New York: Wiley, 1954. Pp. 1–21.
49. SPENCE, K. W., & Beecroft, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48**, 399–403.
50. SPENCE, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.
51. SPENCE, K. W., FARBER, I. E., & TAYLOR, ELAINE. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.
52. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
53. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
54. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–66.
55. WESLEY, ELIZABETH L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.
56. WESTROPE, MARTHA R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515–524.
57. WHITTAKER, EDNA M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733.

Received July 1, 1954

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 4, 1955

LEARNING DURING SLEEP?

CHARLES W. SIMON AND WILLIAM H. EMMONS
The RAND Corporation

Approximately 22 years of the average man's life is lost in sleeping. Economically minded persons and harassed students have long searched for some means to use this time to further advantage. For nearly a third of a century, now, there has been a growing interest in the possibility of trying to learn while one sleeps.

The science fiction writers were among the first to propose sleep-learning as an educational technique. In 1911, a magazine called *Modern Electrics* published a fiction story by Hugo Gernsback (**4**) in which the hero learned during sleep. The process was simple. Upon retiring, the student placed the material to be learned on a machine which played it automatically while he slept. In 1932, Aldous Huxley (**9**), in his *Brave New World*, scoffed at the possibility of improving the *S*'s intellect while he slept, but described the world of the future as one which utilized the sleep period to train the lifelong attitudes of its populations. Other science fiction writers have continued to use the idea of sleep-training in their stories.

Popular news and picture magazines, citing experimental evidence, constantly reinforce the public's interest and misinformation about this osmotic form of education. From time to time, a national news service will release a story telling of some individual who works for a living by day and studies chemistry, operatic arias, or other lessons while sleeping at night. On "People Are Funny," a popular radio stunt show, an audience participant learned a few short sentences and a girl's name while he was apparently asleep on the stage before the audience.

Following closely on the heels of public interest have come the entrepreneurs. Commercial firms have sprung up throughout the country selling recording devices which automatically turn on and off to permit the individual to learn while he sleeps. Impressive publications have been distributed by many of these firms propagandizing their products and extolling the validity and value of sleep-training (**19**, **21**). On the whole, their claims of success for sleep-training have been extrapolations beyond the available supporting data. The majority of claims have been based on distorted facts, statements by unqualified authorities, and armchair hypotheses. Noncritical use of the information, anecdotal evidence, and the citing of inadequate research, much of which will be reviewed in the present paper, have made these commercial publications poor criteria for the support of the validity of the sleep-learning process.

In its constant search for new methods to speed training during national emergencies, the military, too, has tried to sleep-train its personnel. In both the first and second World War, exploratory efforts were made[1,2] to teach service men Morse code during their sleep. A VA Hospital doctor reported the successful use of sleep-training as a supplement to waking-training in a mental health

---

[1] Comdr. R. R. Humes, USN. Personal communication. January 25, 1954.

[2] L. L. Thurstone. Personal communication. October 19, 1953.

program (**20**). Unfortunately, most of these incidents have been too inadequately controlled to judge their effectiveness during sleep, and in general their results have not been sufficiently striking to justify the extra energy and expense required to carry them out.

It is the scientist who has been slow in investigating these claims for successful sleep-learning. Although psychologists have been concerned with problems of learning from their formalized beginning, there has been relatively little interest in the psychological processes which take place during sleep and a rather skeptical—although untested—attitude toward the value of studying sleep-learning. One of Pavlov's students, Krasnogorski (**11**), did attempt without success to condition the salivary reflex in a young child during sleep. Recently, however, a resurgence of interest in the problem has occurred. In the present paper the available research on the sleep-learning phenomenon has been collected and subjected to a critical evaluation.

Ten such studies were found. Of these only three have been published (**3, 12, 13**). Five were academic theses at the master's or bachelor's level (**1, 2, 5, 8, 23**) of which one was read at an APA meeting (**23**). The remaining two[2,3] were described in private communications from the authors. Although a number of other references were found, insufficient details were available for an adequate review or else they could not be classified as research under even the broadest definition.

The criticisms in this paper are not for the purpose of belittling these pioneer studies. These criticisms are believed necessary since these studies—many of which are unpublished—are being cited and their conclusions accepted both by the general public and some scientists. Because of the unusual methodological problems in this area of research, a review may help future workers to avoid similar mistakes.

Description of the Sleep-Learning Studies

The experiments reviewed in this paper are described briefly below. In some of the studies, *S*s were eliminated for various reasons. Only the number of *S*s actually completing the experiment is reported.

Thurstone,[2] in 1916, supplemented the waking training of 16 Navy men in a Morse code course with material presented during sleep. The criteria of sleep used in this study are unknown. The *S*s finished the course three weeks earlier than had been expected. Thurstone concluded that the results "indicated some gain" for the sleep-trained group in their ability to send and receive Morse code.

LeShan (**12**), in 1942, tried to break the fingernail-biting habit of 20 boy campers, ages 8 to 14, by playing the phrase, "My fingernails taste terribly bitter," through a loudspeaker 300 times a night for 54 nights. The *S*s were asked if they were awake before turning on the input; if the *S*s appeared restless, the input was turned off. The nails of the *S*s in the experimental group and an equal number of *S*s in a control group were checked every two weeks for eight weeks for signs that the nail-biting had stopped. Since 40 per cent of the experimental group and none of the control group stopped biting their nails, LeShan felt this indicated "the possible therapeutic use of suggestion during sleep."

LeShan[3] in 1943, taught a single *S* a different list of nonsense syllables each morning for 12 days. On the fifth and eighth nights, the list to be learned the next day was repeated 50 times while *S* slept. The criteria of sleep used in this study are unknown. Fifty per cent fewer trials were required to learn the two sleep-trained lists than to learn the ten nonsleep-trained lists. LeShan concluded that the sleep-training facilitated learning.

---

[3] L. LeShan. Personal communication. October 27, 1953.

Elliott (**2**), in 1947, studied the effects of 30 repetitions of a list of 15 common three-letter words during sleep on the per cent saved in learning the list by the anticipation method the following day. Two groups of 20 male college students of equal learning ability spent one night adapting to the laboratory environment. The word list was repeated serially to one group of *S*s while they slept. An EEG record was used as a criterion of sleep. No material was played when the *S* showed "clear alpha patterns" although the EEG was turned off when *E* believed *S* would stay asleep. On awakening the next morning, the two groups learned the sleep-trained list by the anticipation method and were asked if they awoke during the night. The group receiving the sleep-training showed a significantly greater percentage of savings than the control group ($p = .05$).[4] No significant differences between groups were found on the basis of errors or absolute number of trials to learn. Elliot concluded that there is some retention of auditory material during sleep.

Hedges (**5**), in 1950, tried to improve the speech of three mentally retarded and aphasic children (ages 11, $7\frac{1}{2}$, and 6) by sleep-learning. Short sentences, paragraphs, and simple consonant sounds were played to the children between 433 and 1501 times distributed over 7 to 13 nights. The machine was not turned off if an *S* awoke. The third child required fewer waking lessons to learn the sleep-trained consonant than it had required previously to learn a nonsleep-trained consonant. The babbling of the first child increased. The second child showed no effects. Hedges interpreted this as showing "the possibility of perhaps speeding" the training of the third *S*.

Fox and Robbins (**3**), in 1952, taught 30 college males and females a list of 25 pairs of English-Chinese vocabulary. The *S*s were divided into three groups, matched on a pretest. Two vocabulary lists were prepared; in the first list, word pairs were matched as they were to be learned on the posttest (the facilitation list), while in the second, they were mismatched (the interference list). These lists were played 15 times to two of the *S* groups approximately three hours after they had retired to go to sleep. The third group—a control—heard only music while they slept. Although *E*s did not observe *S*s during the input period, they disqualified all *S*s who said they had awakened during the night. The facilitation group required fewer trials and the interference group required more trials to learn the posttraining list than did the control ($p < .001$). The *E*s concluded that learning could occur during sleep and could be detected by the savings method.

Leuba and Bateman (**13**), in 1952, believed that they had taught the words of three songs, 8 to 27 lines long, to a lightly sleeping *S*. Each song, unknown to *S* beforehand, was played to her five times a night for three successive nights as she slept. During the input period, *S* was not observed by *E*, although for three brief periods, *S*'s husband observed no restlessness on the part of the *S*. The *S* claimed she did not awaken during the night. Given only the song title the next day, *S* was able to write the lyrics of two out of the three songs with no errors and of one with only minor errors. The *E*s believed the sleep-training was successful. No learning occurred in a later study during and following the use of sedatives.

Hoyt (**8**), in 1953, taught ten pairs of English-Chinese vocabulary to his *S*s. Twenty *S*s were matched on a pretest and given one night to acclimate to sleeping in the laboratory. On the experimental night, one group of eight *S*s received twelve repetitions of a list paired in the same manner as that to be learned the next morning (facilitation list). A second group of eight *S*s received a list of comparable words, but with the pairs mismatched (interference list). The remaining four *S*s acted as a control group and received no sleep-training. During the input period, *S*s were observed by *E*.

---

[4] The p values representing the results of tests of significance are all reported as if they were based on a two-tailed t test, irrespective of what the author used. The p levels greater than .15 were not reported.

If they awoke or heard the stimulus material, they were to turn on a light. If this occurred, or if they stirred, the recording was turned off until *S* lay quiet for a suitable length of time. The next morning *S*s were asked if they had awakened during the night. Those that had not were given the paired vocabulary list to learn to two correct anticipations. Statistically insignificant differences were found to favor greater savings for the interference group rather than the facilitation group in both the mean number of trials required to learn the list as well as the number of correct responses occurring before the criterion was reached. No comparisons were made with the control group. Hoyt felt that under the conditions of this study, learning during sleep could not be detected.

Stampfl (**23**), in 1953, had six college males learn lists of ten nonsense syllables while acting as their own controls. Different lists were repeated 4, 8, 16, and 32 times on different nights while *S*s slept, and were learned to one correct anticipation on the morning following the repetition of a particular list. At other times, the same *S*s were tested on other lists after no sleep-training. Before presenting the stimulus material, *S*s were asked if they were awake. During the input period they were watched, and if movements occurred, the input was stopped. Since no significant differences were found for savings in either trials or errors between learning sleep-trained and non-sleep-trained lists, no comparisons were made among performances after different numbers of training repetitions. Stampfl felt that the sleep-learning hypothesis was uncertain and improbable.

Coyne (**1**), in 1953, carried out a series of exploratory studies on a variety of psychological problems using from four to six male college students as *S*s. He believed the results to be generally favorable for sleep-learning. Unfortunately, a statistical error was discovered in Coyne's thesis after he had written his discussion and conclusions which led to a more favorable interpretation of the results than was justified. Although *E* could observe *S* as he slept, he primarily depended on *S* to press a buzzer as soon as he awakened during the night. The problems Coyne studied are described below.

1. An interference list of adjectives was presented 25 times to the sleeping *S*. A similar list was learned to one perfect anticipation the next morning. A control group had no sleep-training. The sleep-trained group did poorer than the control group ($p=.07$).

2. When the first problem was repeated using *S* as his own control, more errors were made on the list which had received the interfering sleep-training ($p=.15$). No difference in savings (trials to learn the posttraining list) was observed.

3. Twenty pairs of numbers and words were presented 24 times during a single 45-minute period to the sleeping *S*s. These same *S*s also received another list for the same number of repetitions distributed over a four-hour period. On subsequent mornings, *S*s were required to answer the appropriate word when given the number found on the list on which they were trained during the night, after that list had been mixed with ten additional new number-word pairs. A greater percentage of sleep-trained words were associated with the correct numbers than those words not trained during sleep ($p=.15$). Distributed sleep-learning resulted in fewer recall errors than did massed sleep-learning ($p=.05$).

4. While the *S*s slept, one list of adjectives was repeated 100 times; on a different night, another list was repeated 25 times. On the morning following the sleep-training, the list was learned to one perfect repetition. No significant differences were found in performance between the two amounts of training.

5. The *S*s were required to solve a number of problems by finding *E*'s solution of a concept composed of the correct combination of letters, colors, and their relative positions. The solution to one problem had been repeated 180 times while *S* slept the night before. Insignificant differences were found favoring the performance on the sleep-trained problem.

6. While *S*s slept, one list of nonsense syllables was presented 30 times begin-

ning at two o'clock and, on a different night, another list was presented 30 times beginning at five o'clock. Since it was discovered that the two lists were of unequal difficulty, comparisons between performance at different times of presentation were not made.

7. During sleep, the contents of a particular picture were described to *S*s 90 times. The following morning, a number of out-of-focus pictures were shown, including the one described during sleep. The degree of focusing required before *S* could identify the pictures was the measure of learning. No significant differences in the case of identifying sleep-trained and nonsleep-trained pictures were found either when *S*s were asked to give unaided responses or when multiple-choice solutions were provided.

Rather than examine each of the ten sleep-learning studies independently, this paper has been organized to discuss them all on the basis of the following categories: experimental design, statistical considerations, methodological considerations, and sleep criteria.

EXPERIMENTAL DESIGN

The use of a control group or of using *S* as his own control is recognized as a necessary procedure in order to know whether a certain experimental effect is real or not. In a number of the sleep-learning studies, however, both of these techniques were conspicuously absent or inadequately handled.

Thurstone (see footnote 2) recognized the inadequacy of his *uncontrolled experiment* teaching Navy men Morse code and attempted to run a second study in order to compare the performance of one group which received sleep-training with another which did not. This study was discontinued, he reported, when it was discovered that ambitious Navy instructors of the control group had been giving extra daytime instruction in order not to be out-taught.

Hedges (**5**) used *no controls* with two of the three speech-defective children he attempted to sleep-train. Since improvement in speech is a maturational problem, even for retarded children, the need for a control was paramount. The increased babbling of the first child may have been independent of the sleep-training and due solely to an additional month or two of growth, although Hedges believed that it took place immediately after the introduction of the sleep-training. Since *S* also received waking-training on the same material, it is impossible to know to what extent one can attribute the increase in babbling to sleep-training. No apparent learning took place with the second *S*. Although Hedges' third *S* acted as his own control by learning to pronounce one consonant with sleep-training added to his waking-training and another consonant without the sleep-training, this was an inadequate control measure since the experimental design was such that the sleep-trained consonant followed the nonsleep-trained consonant. Any improvement in the latter could be attributed to practice as well as to maturation.

Nor was a control used in the work of Leuba and Bateman (**13**). In this study, *S* presumably had no previous knowledge of the songs played to her during sleep, yet was able to write the lyrics without further training on awakening. If this were the case, any learning which took place during sleep would be a significant improvement, although materials such as songs and poetry probably have some internal predictability. This form of experimental design represents an *implied control*. It is implied that *S* acted as her own control for had she been tested previously, no appropriate responses could have been made, nor could one suspect that maturational factors were operating to produce positive results.

A number of *E*s actually used their *S*s *as their own control.* In LeShan's (footnote 3) second study, he compared the number of trials an *S* required to learn a list with and without sleep-training over a period of days. By having the nontrained periods before and after the sleep-trained periods, the superiority of the sleep-trained lists could not be attributed simply to practice or maturation. On many of his studies, Coyne (**1**) gave sleep-training to his *S*s one night and no sleep-training on the next, counterbalancing the order for these procedures between two groups of *S*s. However, when the *S* acts as his own control, either the study material which is used under the varying conditions must be carefully equated, or additional counterbalancing must be introduced into the experimental design to correct for the inequalities. No *E* used this counterbalanced design; some made use of the published tables on which similarities and associability of the stimulus material had been previously calculated for the items. Several of Coyne's (**1**) studies were unanalyzable after he discovered the study material had not been equated.

Four of the experiments were designed to use *separate control groups.* In Elliott's (**2**), Fox and Robbins' (**3**), and Hoyt's (**8**) studies pretesting took place in order to equate the mean performance of the control and experimental groups. LeShan (**12**) used unmatched groups in his study with fingernail-biting children. The median age of his twenty experimental nailbiters was slightly less than ten years. His control was divided into two groups consisting of eight nailbiters with a median age of nine years, and twelve more nailbiters with a median age of twelve. Some question might be raised concerning LeShan's failure to better equate the experimental and control groups on age, for there is reason to suspect a relation between nail biting and age. Wechsler (**25**) found a sharp rise in nail biting for boys around the age of twelve; if this is so, there would be a lower probability for the older control group to stop biting its nails than the younger experimental group, thus reducing the effectiveness of the older group as a control.

Two of the experimenters added a *second experimental group* to their design. Fox and Robbins (**3**) and Hoyt (**8**) used both a facilitation and an inhibition group in order to obtain a more sensitive indication of the value of sleep-training. Of all of the studies, only that of Fox and Robbins (**3**) provided the control group with a neutral stimulus—music—for the same amount of time as that in which the experimental group received the verbal test material. If there are any disturbances during sleep due to the stimulus and if these in turn affect recall, such a procedure is a wise one.

Although none of the *E*s were directly concerned with the problem, the use of an *additional control group* to compare results from sleep-training with the equivalent amount of waking-training would have been quite illuminating.

STATISTICAL CONSIDERATIONS

Only five of the *E*s (**1**, **2**, **3**, **8**, **23**) treated their data statistically to see if the sleep-training improved performance significantly. The remaining *E*s used clinical criteria to evaluate the effects of sleep-training.

Although Elliott's (**2**) results favored the performance of the sleep-trained group over the nonsleep-trained group, he failed to find a significant difference at the 5 per cent level in the number of trials it took to relearn the training list. Elliott had equated his groups on a pretest, but did not attempt to match the individual *S*s for the analysis. Since equating tends to decrease differences between means, failure to remove the variance due to *S*s inflates the error variance and decreases the probability of getting significant differences. When the present authors (**22**) did an analysis of covariance with pre- and posttest scores from Elliott's data, the differences between the mean number of trials to learn a new list by sleep-trained and non-sleep-trained *S*s were significant below the .05 probability level.

Coyne (**1**) used the *one-tailed t test* to evaluate his data. The wisdom of this treatment is questionable. In order to avoid abuses and controversy, exploratory work should be as cautious in its interpretations of results as it should be daring in attempting new ideas. Using the one-tailed *t* test does not allow for the possibility that differences might be in the direction opposite to that hypothesized (**6**). This is serious in any exploratory work; it is particularly dangerous in sleep-learning studies where one could seriously suspect that the intervening training during sleep might actually hinder normal waking recall.

Half of the *E*s (**2**, **3**, **8**, **12**; footnote 2) used a reasonably large *number of Ss* as compared to the number used in most psychological studies; in the remaining cases, the number was smaller. When the number of *S*s is small, one might be more critical of accepting the null hypothesis merely because the level of significance was not below the traditional 5 per cent level. For studies as exploratory as these, significance levels of 15 per cent could be arbitrarily considered encouraging. Since the expense in time and money is relatively small, preliminary work in sleep-training should favor fewer Type II errors in order not to reject valuable experimental leads by accepting a false null hypothesis.

METHODOLOGICAL CONSIDERATIONS

A number of *E*s believed that differences in methodology might have been responsible for the divergent successes and failures found in the sleep-learning studies. These and other considerations are discussed below.

Subjects

The majority of the studies employed the traditional college student—male and female—as *S*s. Thurstone used Navy men and LeShan used young boy campers. Hedges (**5**) bravely attempted to provide sleep-training as a supplement to the waking-training of children who had speech deficiencies and who were suspected of being mentally retarded. None of the *E*s attempted to study the effects of either age or sex on sleep-learning.

The selection of *S*s may have an effect on whether successful sleep-training results are attained or not. Underwood, while reviewing Fox and Robbins' paper in the *1953 Annual Review of Psychology*, commented that "such low variability [on performance scores] among *S*s on the test list is rarely found in normal transfer experiments with such material, but this may again only reflect the presence of a highly select and homogeneous sample" (**24**, p. 48). Low within-group variability was certainly in part responsible for the high degree of significance of the differences between the sleep-trained groups and the control.

Whatever the variability of the group, it would appear wiser to choose individuals who have shown the capacity to learn while awake. Perhaps the effects of sleep-training are so subtle that its benefits will be found only when it is applied to individuals with very high IQ's.

Number of Repetitions

The experiments can be divided into two groups on the basis of this variable—those who gave an exceptionally large number of repetitions of the material during sleep (**5,12**) and those that gave significantly fewer repetitions (**1, 2, 3, 8, 13, 23**; footnote 3). The large number of repetitions were actually spread over a number of nights and ranged from a total of 433 times over a period of eight nights to 16,200 times over a period of 54 nights. The smaller number of repetitions ranged from 8 to 180 times, with one study (**13**) playing the material five times per night for three successive nights. More repetitions were characteristic of the field as opposed to the laboratory studies. Both groups obtained both positive and negative results, although the tasks were varied sufficiently so that direct comparisons could not be made. LeShan's (**12**) study, in which 40 per cent of the experimental group stopped biting its nails, represented a successful example where a great deal of repetitious training seemed to have affected a semi-involuntary behavior.

Stampfl (**23**) and Coyne (**1**) found no differences when they tried to study the effects of different amounts of repetitions on sleep-learning. Actually since neither *E* used a non-sleep-trained group as a control, neither could conclude that any sleep-learning took place at all in this phase of their study. Coyne (**1**) suggested that in his study there may have been no greater differences in savings after one hour of repetitions than after four hours because the material was simple enough to be learned in one hour and additional practice could not improve it. This is a reasonable hypothesis, as is its antithesis—that during sleep, learning is sufficiently slow so that little is learned even after four hours of repetitions. Sleep-learning, if it is to occur at all, may require that an extremely large number of stimulus repetitions take place. It will be important to evaluate sleep- versus waking-training from the standpoint of economy of both time and effort.

Presentation

The manner in which the study material is presented to the sleeping *S* has been considered by some *E*s as critical to the success or failure of the training. Two dimensions of this variable are the time of presentation and the order of presentation.

The problem of *presentation time* is an important one since during the sleeping period, time is related to some extent to the depth of sleep, which in turn may be related to trainability. Deeper sleep tends to be more prevalent in the early period of sleep, while lighter sleep tends to occur later (**18**). Of course, the levels vary considerably throughout a normal night's sleep.

Coyne (**1**), studying the effects of presenting the material at different times during the sleep period, failed to equate his lists beforehand and could draw no conclusions from his results. He recognized that presentation time might be inversely related to the amount retained. He suggested, however, that this was due to the recency of the presentation to the recall period. It is interesting to speculate whether or not the Jenkins-Dallenbach interference effect (**10**) occurs within the sleep period for materials presented during sleep.

The differences Coyne found between massed and distributed learning might also be accounted for on the basis of presentation time. Although he concluded that distributed sleep-learning was superior to massed sleep-learning, there is no way to determine whether this was so because of the spacing of the training or because some of the distributed inputs occurred during the period just before waking—often a light drowsy state—while the massed inputs occurred only during the deeper and possibly less receptive period.

The *order* in which the material was presented during sleep may also affect the results of sleep-training. In a waking state, serial learning has been shown to be easier than learning material presented in a varying order (**7**). Thus, if any learning takes place in sleep, a serial presentation would more probably increase any positive effects which might occur. Also, if the sleep state is one in which no mental organization takes place, this would favor the learning of only the more organized serial presentation.

Hoyt (**8**) and Stampfl (**23**) varied the order in which their material was presented and failed to find that any sleep-learning took place. Fox (**3**), LeShan (**12**), Elliott (**2**), and Leuba (**13**) believed they found evidence of successful sleep-learning. Thurstone probably varied the order of his material, but his *S*s were practiced over a period of months so that the effects of presenting the material in a varying or an unvarying order may have been minimized.

The varying presentation order in Hoyt's study was methodologically different in one major respect from that used in Stampfl's study. Since the former study used paired-associate material, varying the order in which the pairs were presented would have less effect than it had in the latter study where the order of words in a list was varied during sleep-training, even though they had to be recalled serially during the waking period. Any positive effects of sleep-training in the latter case may have been neutralized by the negative effects built up through nonserial learning.

Training Problems, Materials, and Mode of Input

The types of psychological problems studied by the majority of sleep-training investigators have been quite limited. With the exception of Coyne's (**1**) work on concept formation and perceptual sets and LeShan's (**12**) and Hedges' (**5**) therapeutic studies, the remainder of the research has been involved with training problems which require the memorization of word lists. It is unlikely that all types of problems are suitable for sleep-training, although exploratory studies should examine many rather than a few possibilities.

Of the material used in the studies requiring verbatim recall, the degree of meaningfulness ranged from lists of nonsense syllables through short words to foreign language vocabulary. We know that the more meaningful the material, the easier it is to learn in the waking state (**7**). Stampfl (**23**) suggested that this might explain why his results with nonsense syllables were poorer than Elliott's (**2**) who used a list of adjectives, and why Fox (**3**), using a Chinese-English vocabulary, got even more striking results. This does not seem to be a critical variable, however, for Hoyt (**8**), using the same Chinese-English vocabulary as Fox (**3**), got negative results, while LeShan (footnote 3) and Coyne (**1**), using nonsense syllables, got positive results.

All of the *E*s used an auditory input. This is certainly the most obvious technique and would appear to be the most economical; however, other sensory channels need not be ignored. It may be necessary to flash lights on closed eyelids or to apply tactual stimuli to the fingertips in code in order to "reach" the sleeping *S*.

Techniques and Measures of Retention

The techniques used to measure the retention of training material can have considerable influence on the amount of material recalled. However, in the present studies, there did not appear to be a pattern of successes or failures consistent with the technique used.

A number of *E*s did not require the verbatim recall of the verbal material presented during sleep. Instead, they evaluated retention on *S*'s ability to *use the material* in posttraining tasks or on an *observed change* in *S*'s behavior after the training. LeShan (**12**) examined his *S*s' fingernails every two weeks to see if the children had stopped biting their nails. This technique might have been slightly more objective had the examiner not known which children were and were not receiving the sleep-training. Hedges' (**5**) clinical evaluations of his children's speech improvement required even more subjective judgments. Because of the very complexity of this measure, Hedges wondered whether it was sensitive enough to detect improvement. Coyne (**1**), in several studies, had his *S*s describe an out-of-focus picture, the identity of which had been given to them during sleep, and to determine *E*'s solution of concepts composed of the correct combination of certain stimulus variables, the answer to which had also been provided during sleep. Thurstone's *S*s were evaluated on their ability to send and receive Morse code. Both Coyne's and Thurstone's results could be quantified.

Leuba (**13**) gave his *S*s the titles of songs played during sleep and required them to *recall unaided* the lyrics. Positive results were claimed for sleep-training in this study.

The *savings method* has been used by a majority of *E*s (**1**, **3**, **8**, **23**; footnote 3), for they believed it to be a more sensitive measure of retention. Stampfl (**23**) believed that although material could not be consciously recalled at any moment, its presentation during sleep may have still modified the nervous system sufficiently to make learning easier and a savings effected. However, use of the savings technique may confuse the measure of retention with a measure of the ability to learn since both are confounded within the same performance score. None of the *E*s using the savings measure compared the time to relearn a list with the time it took the same *S* to learn an equivalent list presented after sleep. Comparisons with the original list presented before sleep are not sufficient. The positive transfer which occurs from the pretest to the posttest—the phenomenon of "learning how to learn"—may account for some of the apparent savings which several researchers believed they found. However, Stampfl (**23**) gave three practice lists to be learned before the experiment in order to bring *S*s' learning curve closer to its asymptote and thus reduce the "learning to learn" effect.

A still more sensitive technique for measuring retention is that of *recognition;* at least in Luh's (**16**) classical study this was so after a two-day interval between training and the posttraining test. Coyne (**1**) used this technique in the form of a multiple-choice test for one of his exploratory studies on perceptual set, but still failed to get positive results. Hoyt (**8**) told of an exploratory study in which he presented a single number-word pair to a sleeping *S* who was under constant observation for a total of 11 hours on three successive nights. On awakening after the third night, *S* was given the number and asked to pick the correct word from a group of ten. He was unable to do this correctly. A major difficulty with the recognition method is that it must be corrected for chance guessing. With words, however, this correction for chance is difficult since all words do not have an equal probability of being recognized (**17**).

Variations in the score used to measure retention have failed to consistently differentiate success and failure in sleep-learning. Many of the *E*s (**1**, **2**, **8**, **23**; footnotes 2, 3) used the *number of trials* to learn posttraining material to one perfect repetition as a measure of retention. Hoyt (**8**) required that the lists be learned to two perfect anticipations; this may tend to make his results slightly more reliable. Some *E*s (**1**, **2**, **8**, **23**) also used as measures of retention the *number of errors* or correct responses made on the first trial or later trials, or on cumulative trials. Where the *S*s performance was used to measure retention, the measure was characteristic of the task, e.g., Coyne used the extent of focus of the projector as the measure of performance in his perceptual set study. The measure that is used may determine in part the results which are found. For example, Coyne (**1**) noted that the measure by trials often favored the opposite results than those favored by the measure by errors. Error measures generally lead to less conclusive results than those measured by trials. This situation does occur in waking-research and need not be contradictory, although it certainly affects the conclusions drawn as well as the practical applications of the results.

Before final conclusions can be drawn concerning the feasibility of sleep-learning, more recall techniques should be tried. Simon and Emmons (**22**) discuss this while considering the possibility of secondary cerebral storage mechanisms for material introduced during sleep.[5]

SLEEP CRITERIA

Perhaps the most damaging criticisms of the sleep-learning studies to date have been the inadequate control of sleep and the criteria used for defining sleep. So elusive is the process of sleep from the psychological standpoint and so little is known about the actual physiological mechanisms which cause sleep that the problem of determining whether *S* is asleep is to a great extent a semantic one. Although both direct observation and *S*'s subjective report are reasonably reliable for deciding if *S* has been asleep over a block of time, neither can be considered sufficient to know if *S is* asleep at any moment in time. Therefore, the care with which *E* determines the sleep condition of his *S* at the time of the input is highly important and determines the degree of confidence which one may place in the conclusion that learning during real sleep is or is not possible. Some *E*s (**1**, **2**, **3**) eliminated

---

[5] Dr. Bernard Fox, in a private communication (26 June 1954), described his attempt to use hypnosis as a means of facilitating the recall of sleep-trained material. A 55-year-old man was brought to a point where deep hypnosis was possible. In his normal sleep he was presented with a list of ten words repeated for a half-hour. During this period, *S* was observed and the input turned off whenever he moved. The next morning, he was unable to recall any words unaided. When he was hypnotized, he made only one association which *might* have been related to one of the words. Still under hypnosis, he recognized fewer words than would be expected by chance when the ten were interspersed randomly in a group of thirty. Neither reading the words aloud nor making him choose by a forced-choice method produced any more positive results.

S from the experiment if he awoke or said he heard the stimulus. Other *E*s (**12, 23**) shut off the stimulus until *S* went back to sleep.

In four of the studies (**2, 3, 8, 13**), *E*s *asked* Ss *after the training if they had awakened*. This is an unsatisfactory criterion of sleep for experimental studies since it is a common experience to awaken during the night, perform a number of rational acts before retiring again and remember nothing the next day (**8, 18**). Hoyt stated that one *S* was observed to awaken during the night, leave the laboratory to go to the bathroom, and later return to bed. The next morning, *S* remembered nothing about it. In two of these studies (**3**; **13**) *E*s were not even present to *observe* Ss as an added check.[6]

In one experiment (**3**) the *time of night* was considered a partial criterion of sleep; training was applied during the period when studies have shown sleep to be at its deepest. However, published sleep curves are based on statistical averages and cannot be used to predict accurately the sleep characteristics of the single individual. Individual sleep curves vary considerably from time to time during the night (**15**).

Two of the *E*s (**12, 23**), *before presenting the stimulus material, asked Ss if they were asleep*. Since LeShan's (**12**) *S*s were adolescent boys in camp, naive as to the purpose of such questioning, one may wonder if asking them if they were asleep would be a sufficient check. It has been observed that even willing *S*s often fail to respond to such questions, although they admit hearing them later. Another major difficulty with this technique is that *S* may actually have been asleep when the test question was asked, but later would awaken sufficiently to hear the subsequent input. In these studies, no systematic check was made to be sure that *S*s did not awaken *after* the initial test, except to note if they moved.

Coyne (**1**) and Hoyt (**8**) had their *S*s *press a button whenever they awoke* in order to stop the input of training material. This method alone is not sufficient since it has been found that often the desire as well as the ability to do this is not present although *S*s are able consciously to hear and understand the material. Since much of the material used in these studies was short and specific, even a few seconds of waking presentation would have been sufficient to allow *S*s to hear and learn much of the stimulus while awake.

Stampfl (**23**), LeShan (**12**), and Hoyt (**8**), in addition to asking their *S*s if they were or had been awake, also *noted whenever they moved*, and turned off the machine at that time. This criterion of sleep is not completely satisfactory since EEG records indicate that sleep may lighten to almost a wakening state before movement occurs, as well as afterwards, or when there has been no movement at all (**15**). Hoyt (**8**) found that two of his *S*s said they had awakened but showed no observable movements, while three subjects said they had not awakened but showed from two to six movements each. This suggests that the criteria of asking and noting movements did not correlate very well. In Hedges' (**5**) study, the parents occasionally observed the children to see whether or not they awoke during the sleep-training; two children awoke. In the first case the number of training repetitions were not counted for the time *S* was awake. However, without a waking control, this correction is meaningless. The other *S* awoke while the recording was playing and began drowsily to follow the instructions given on the record. Hedges felt this aided the progress observed in that case.

---

[6] Fox stated "if any *S*s in the group of 30 actually did wake up without reporting it, then the number was probably distributed about equally throughout the groups. Moreover, since all *S*s with one exception show effects in the expected directions, it is clear that the results cannot be explained on the bases of a few *S*s in each group who awakened but failed to report doing so" (**3**, p. 78). It may be on the basis of many.

Elliott (**2**) used the *electroencephalogram* to determine the *S*'s sleep level. Independent workers using a variety of techniques have established a significant correlation between the brain wave patterns and the depth of sleep (**14**). Because the writers believe that properly evaluated electroencephalograms represent the most objective, continuous, and practical on-the-spot indicator of sleep depth available, Elliott's (**2**) positive results—often quoted by the commercial firms in support of their claims for sleep-learning—are quite noteworthy. An examination of his sleep criterion revealed one major flaw, i.e., Elliott did not keep the EEG running during the entire training period. Therefore, no continuous check was made of the *S*'s sleep level while the stimulus was being played. The EEG was turned on at the beginning of the training and kept on until *E* believed that sleep would remain deep enough; then he would shut it off for the remainder of the half-hour stimulus period. With Elliott's data, a correlation of −.39 was found between the amount learned and the amount of monitoring. This was not significant at the 5 per cent level but in a direction which suggested that more savings occurred with less monitoring. The writers (**22**) also compared the amount of savings made by a group of Elliott's five *S*s who were monitored nearly all of the training time with the savings of the remaining fifteen *S*s who were monitored on an average slightly more than a fourth of the time, and never more than 54 per cent of the time. For the first group, the average savings in trials to learn was 27.5 per cent; for the second group, the average savings was 45.0 per cent. The difference of 17.5 per cent was highly significant below the 2 per cent level of significance and we can conclude that when a thorough check of the sleep condition was made with the EEG during training, considerably less savings took place.

That any savings occurred for the groups monitored nearly 100 per cent of the time might be accounted for in two ways without assuming that learning actually occurred during sleep. First, the apparent savings may be simply an effect of the "learning how to learn" phenomenon discussed earlier in the section on "Methodological Considerations." Second, Elliott stated that whenever an alpha pattern was "clearly" present he concluded *S* was awake and turned off the input machine. In so doing, he quite likely permitted many of his *S*s to listen while they were already awake since movement, tenseness, or the opening of the eyes may sometimes block a clearly established alpha.

The expense of the EEG equipment negates its widespread use. This does not mean that without it, however, adequate sleep-learning studies cannot be done. The use of a combination of the criteria and continuous monitoring may be sufficient to insure some rough but adequate control of *S*'s sleeping condition. It is interesting to note that Hoyt (**8**) and Stampfl (**23**), finding negative results, were the two *E*s using the greatest number of multiple and continuous criteria.

The continuity of monitoring cannot be overemphasized; for in spite of wishful thinking to the contrary, *S*s do awaken while the input is being presented. Two out of three of Hedges' (**5**) *S*s awoke. As many as half of Coyne's (**1**) *S*s awoke in some of his studies. Elliott (**2**) reported that six of his *S*s awoke. Only two out of sixteen of Hoyt's (**8**) *S*s in the experimental group failed to awaken during an input period. LeShan (**12**) also stated his *S*s awoke, but did not specify how many. One-fourth of Fox's (**3**) *S*s awoke, and were eliminated from the study.

The conditions under which *S*s sleep can influence to some extent the soundness of their sleep. Elliott (**2**) and Hoyt (**8**) required *S*s to sleep in the laboratory one night previous to the experiment. Any sleeplessness occurring on this first night probably induced a deeper sleep on the second. Hedges (**5**), Fox (**3**), and Leuba (**13**) used *S*s while they slept in their own homes, while Thurstone (footnote 2), LeShan (**12**), and Stampfl (**23**) used *S*s who had been sleeping in familiar "homes away from home" for some time.

Although most of this discussion has centered around the problem, Was the sleep state deep enough?, Stampfl (**23**) and Leuba (**13**) suggested that there is an intermediate point between waking and deep sleep that is optimal for sleep-learning. Some negative results, they believed, may have been obtained because *S* was too *deeply* asleep. Whether this is true or not is an experimental question for future *E*s to answer.

SUMMARY AND CONCLUSIONS

Ten sleep-learning studies were reviewed. Many of these have been cited uncritically by commercial firms or in popular magazine and news articles as evidence in support of the feasibility of learning during sleep. A critical analysis was made of their experimental design, statistics, methodology, and criteria of sleep. All of the studies had weaknesses in one or more of these areas.

It is highly speculative whether or not the studies reviewed in this paper have presented any acceptable evidence that learning during *sleep* is possible. The inadequate control of a number of experimental variables makes the validity of the conclusions drawn by many of the *E*s unwarranted. The conditions under which the results were found tend more to support the contention that some learning takes place in a special kind of waking state wherein *S*s apparently do not remember later on if they had been awake. This may be of great practical importance from the standpoint of economy in study time, but it cannot be construed as *sleep-learning*. More carefully controlled experiments in the future may provide us with a clearer answer to the question," "Can one learn during sleep?," as well as to provide comparative data between waking and resting learning from the standpoint of economy of time and effort. The problem is partially confounded by an inadequate definition of sleep.

REFERENCES

1. COYNE, M. L. Some problems and parameters of sleep learning. Unpublished honors thesis, Wesleyan Coll., Conn., 1953.
2. ELLIOTT, C. R. An experimental study of the retention of auditory material presented during sleep. Unpublished master's thesis, Univer. of N. Carolina, 1947.
3. FOX, B. H., & ROBBINS, J. S. The reten-

tion of material presented during sleep. *J. exp. Psychol.*, 1952, **43**, 75–79.

4. Gernsback, H. Ralph 124C 41+. *Modern Electrics*, 1911, **4**, 165–168.
5. Hedges, T. A. The effect of auditory stimuli presented during the sleep of children with delayed speech. Unpublished master's thesis, Univer. of Wichita, 1950.
6. Hick, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, **59**, 316–318.
7. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.
8. Hoyt, W. G. The effect on learning of auditory material presented during sleep. Unpublished master's thesis, George Washington Univer., 1953.
9. Huxley, A. L. *Brave new world*. New York: Doubleday Doran, 1932.
10. Jenkins, J. G., & Dallenbach, K. M. Obliviscence during sleep. *Amer. J. Psychol.*, 1924, **35**, 605–612.
11. Krasnogorski, N. I. Bedingte und Unbedingte im Kindesalter und ihre Bedeutung für die Klinik. *Ergeb. d. Inner Mediz. und Kinderhk.*, 1931, **39**, 613–730.
12. LeShan, L. The breaking of habit by suggestion during sleep. *J. abnorm. soc. Psychol.*, 1942, **37**, 406–408.
13. Leuba, C., & Bateman, D. Learning during sleep. *Amer. J. Psychol.*, 1952, **65**, 301–302.
14. Lindsley, D. B. Electroencephalography. In J. McV. Hunt (Ed.), *Personality and behavior disorders*. New York: Ronald, 1944, Pp. 1033–1103.
15. Loomis, A. L., Harvey, E. N., & Hobart, G. A. Cerebral states during sleep, as studied by human brain potentials. *J. exp. Psychol.*, 1937, **21**, 127–144.
16. Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**, No. 3.
17. Miller, G. A. Speech and language. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 789–810.
18. Mullin, F. J., Kleitman, N., & Cooperman, N. R. Studies on the physiology of sleep changes in irritability to auditory stimuli during sleep. *J. exp. Psychol.*, 1937, **21**, 88–96.
19. Powers, M. *Mental power through sleep-suggestion*. Los Angeles: Wilshire Book Co., 1949.
20. Schmidhofer, E. Mechanical group therapy. *Science*, 1952, **115**, 120–123.
21. Sherover, M. Learning during sleep: is it possible? New York: Modernophone, Inc., 1954.
22. Simon, C. W., & Emmons, W. H. Considerations for research in a sleep-learning program. Santa Monica, California: Rand Corp. Paper No. 565, 1954.
23. Stampfl, T. The effect of frequency of repetition on the retention of auditory material presented during sleep. Unpublished master's thesis, Loyola Univer., Chicago, 1953.
24. Underwood, B. J. Learning. *Annu. Rev. Psychol.*, 1953, **4**, 31–57.
25. Wechsler, D. The incidence and significance of fingernail biting in children. *Psychoanalytic Rev.*, 1931, **18**, 201–209.

Received July 15, 1954

Psychological Bulletin
Vol. 52, No. 4, 1955

TEACHING IN THE IVORY TOWER, WITH RARELY A STEP OUTSIDE

S. L. PRESSEY
Ohio State University

In recent years, psychologists have literally gone all over the world, and addressed themselves to a great variety of problems. In so doing, they have tried a variety of training and instructional methods, often on the job or in the field, and have evaluated those methods realistically, as with reference to outcomes in production records or combat data. One might hope that all this would have invigorating influences upon methods in the teaching of psychology, and appraisals of them.

Instead, an extensive and careful recent review[1] of investigations of the teaching of the subject since 1942 shows the largest single group dealing with phases of the old comparison of lecture versus discussion in the elementary course, as appraised most commonly by final examinations. And the reviewers seem to excuse the minimal imagination as to methods and minimal results by implying a minimal "degree and nature of change that one can produce in a one-or-two-semester course in psychology" (p. 52). But why are not these limitations of time a challenge to use varied and vigorous methods, chosen on the basis of incisive evaluation? Yet the review seems not to mention any use of laboratory methods, nor any projects, field trips, or field experience. And the great majority of the evaluations seem to have been in terms of paper-and-pencil tests, usually given in the evaluated course (not later or elsewhere). Only talk-talk in a college classroom as a means of bringing about learning, and only markings there on a piece of paper to evidence whether significant learning has occurred!

But many other methods are possible, even in a big elementary course—as the writer knows from over 25 years experience supervising such a course, enrolling some thousand students a year (mostly freshmen), taught mostly by graduate students. There can be informal and social laboratory tasks, ranging from genuine research projects which the students understand as such and are later reported in print, or the making of simple equipment such as charts for the course, to informal half-humorous data gathering regarding the "social psychology" of freshmen. There can be field trips, as to an institution, or field work as helping in the survey of a public school. And evaluations can include careful checking that the laboratory research is well done and the charts well made, notations that the school principal spoke favorably of the survey assistants and that after an institution visit, a girl asked intelligent questions about career opportunities in work there, or an anecdotal record that on a field trip an initial isolate had become an accepted member of a group.

The Wolfle committee doubted the legitimacy of many such doings as either too elementarily useful or too special: one should "teach psychology"; and only certified clinicians or counselors could properly deal with many of these problems. But beginning courses serving mostly students who will not go further in the subject surely should be realistic about them and their needs—can thus best serve them *and* give them some functioning understanding of the subject. And bringing it about that an isolated student gets acquainted need require little more of a good instructor than being a good host; even aid in early consideration of a vocational choice sometimes requires little more than mature good sense, with referral.

---

[1] Birney, R., & McKeachie, W. The teaching of psychology: a survey of research since 1942. *Psychol. Bull.*, 1955, **52**, 51–68.

The reviewers report the extraordinary finding that sophomores get better marks in a first course in psychology than do freshmen, and seem to agree with the inference that the first course should therefore not be given in the first year. But if the first course were not taken till graduate school, grades would be still better! Surely not where students do best is the place for a course, but where they need it the most.

The rather unhappy reviewers conclude with the statement that the most serious need, in research in teaching, is to get it "related in some way to general psychological theory" (p. 64).[2] Climb higher in the ivory tower! Why not instead come down and open the door, watch what students do outside class, see what psychologists are doing in the wide world, and even venture off campus occasionally, to mix with the folks on Main Street.

Received March 24, 1955

---

[2] There *could* well be more consideration of a broad theory of higher education. More than any other subject field, psychology *should* have appreciation of the great variety of gains —in abilities, understandings, personality traits—possible during the college years. Psychologists have a great variety of contributions to make, in furtherance of those gains. Some are highly special, as in counseling or therapy. But the basic contention of this commentary is that there remains a great range of possible contributions, most best attempted by broadened concepts of and methods in teaching—and that the excellent studies summarized in the review cover that range only from A perhaps through C. The plea is for a wider range of experiment.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 4, 1955

ADDITIONS TO PSYCHOLOGICAL NECROLOGY: JAPAN, 1928–1952

KOJI SATO
Kyoto University

In the former Psychological Necrology (1903–1927) by E. G. Boring (2) two Japanese psychologists were listed, namely, Y. Motora and T. Nakashima, but in the new one by Bennett and Boring (1) none were included. Perhaps this is because communication between American and Japanese psychology was interrupted during World War II. However, Japanese psychology has developed a great deal during this period. Among the deceased psychologists in Japan there are a number who are worthy of being included in Boring's new list. From a total of 45 deceased psychologists, I selected 25 names and asked the leading psychologists in Japan, Tatsuro Yatabe (Kyoto Univer.), Toru Watanabe (Nihon Univer.), Matsusaburo Yokoyama (Keio Univer.), Sadaji Takagi (Tokyo Woman's Christian College), Megumi Imada (Kanseigakuin Univer.), Yoshikazu Ohwaki (Tohoku Univer.), and Torao Obonai (Tokyo Univer. of Education), to check them against the level of Boring's list. The following 13 were selected:

Fukutomi, Ichiro, Kenkoku Univer., Manchuria, b. 8 June 1891, d. 27 Feb. 1946.

Hayami, Hiroshi, Keijo Univer., Korea, b. 23 Oct. 1876, d. 27 June 1943.

Iwai, Katsujiro, Kyoto Univer., b. 17 Feb. 1886, d. 2 Nov. 1937.

Kakise, Hikozo, Ministry of Education, Tokyo, b. 13 Sept. 1874, d. 23 April, 1944.

Kuroda, Ryo, Keijo Univer. and Gifu City, b. 30 Jan. 1890, d. 5 Jan. 1947.

Kubo, Yoshihide, Hiroshima Bunrika Univer., b. 22 Apr. 1883, d. 12 Nov. 1942.

Masuda, Koreshige, Tokyo Univer., b. 29 Dec. 1883, d. 7 Aug. 1933.

Matsumoto, Matataro, National Academy of Science, Tokyo, b. 15 Sept. 1865, d. 24 Dec. 1943.

Morita, Shoma, Jikeikai Medical School, Tokyo, b. 18 Jan. 1874, d. 12 Apr. 1938.

Nakamura, Katsumi, Tokyo Univer. of Education, b. 7 Feb. 1901, d. 6 Apr. 1952.

Onoshima, Usao, Ministry of Education, Tokyo, b. 16 Mar. 1894, d. 24 Jan. 1941.

Rikimaru, Jien, Taihoku Univer., Formosa, b. 3 Jan. 1890, d. 4 Dec. 1945.

Takashima, Heizaburo, Toyo Univer., Tokyo, b. 1 Oct. 1865, d. 15 Feb. 1946.

REFERENCES

1. BENNETT, SUZANNE, & BORING, E. G. Psychological necrology (1928–1952). *Psychol. Bull.*, 1954, **51**, 75–81.[1]
2. BORING, E. G. Psychological necrology (1903–1927). *Psychol. Bull.*, 1928, **25**, 302–305, 621–625.

Received March 15, 1955.

---

[1] The authors of this article welcome the opportunity to correct one gross error in their list. Narziss Ach was born on 29 October 1871 (not in 1848).

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 4, 1955

SPECIAL REVIEW

HANDBOOK OF SOCIAL PSYCHOLOGY[1]

DANIEL KATZ
University of Michigan

A handbook in any field of knowledge is an ambitious undertaking. When that field is growing so rapidly that changes of major significance take place between the writing of a manuscript and its publication and when the field ramifies widely with ill-defined boundaries, the task is even more formidable. In this perspective the two volumes edited by Lindzey are a magnificent achievement. They replace all existing texts and sourcebooks or any combination of text and sourcebook. They present clear and incisive accounts of the major theoretical positions in social psychology and they give good coverage to the thinking and research on all the traditional topics in the field. Many of these topics are old chapter headings in social psychology (humor and laughter, facial expression, mass phenomena) which more recent texts have omitted. And there is an excellent section of nine chapters on research methods. Moreover, many of the chapters are more than reviews of the literature in their insightful integration or in their suggestions for future research. Though social psychology has been characterized by sectarian trends and an impatience with an historical understanding of its problems, the *Handbook of Social Psychology* should help to counteract these forces. Some of the chapters are still too selective but the over-all pattern is one of real progress. High editorial standards have been maintained and the writing is not only clear, but for a handbook, unusually interesting. Almost any adequate organization of chapters in this field would have problems in finding mutually exclusive areas but the amount of overlapping in the literature reported by the various authors is not great.

The opening chapter by Gordon Allport presenting the historical background of modern social psychology is one of the finest in the two volumes. It is a convincing argument for the importance of the historical approach. It shows the similarity and, sometimes, the virtual identity of contemporary concepts and problems with the ancient history of our discipline and demonstrates how easily some of the blind alleys in our development might have been avoided with more consideration of the thinking and work already available. If its wise precepts were to be followed we might temporarily solve our overcrowded journal situation by seeking for significant research problems rather than following the latest fad in labels and techniques. This chapter is rightly titled "Historical Background" in its concentration on the similarities between current ideas and their early counterparts. What would also be desirable would be a companion chapter on historical development which with the same bold strokes would show how we have moved from the early stages to our present situation. This, however, is not the conventional task of a handbook. The reviewer would make one minor criticism reflecting his own bias. There is too little material on the objective approach represented by such writers as F. Oppenheimer, E. Durkheim, K. Marx, and E. B. Holt.

---

[1] G. Lindzey (Ed.) *Handbook of social Psychology.* Cambridge: Addison-Wesley, 1954. Vol. I (pp. 588+x), Vol. II (pp. 601–1226+x), $8.50 one volume. $15.00 per set.

Of the five chapters devoted to contemporary systematic positions, those on field theory and on Freudian theory are superior presentations in their clarity, organization, and relevance to social psychological problems. M. Deutsch does a remarkable job of exposition in condensing Lewinian metatheory, the formal concepts of field theory, and the research contributions of this school within the confines of a single chapter. His judgment that the experimental contributions of the movement stem from Lewinian metatheory rather than from Lewin's brilliant insight into empirical problems is probably sound. This does not mean, however, that the significant research activities of the school are the outcome of a formal set of theoretical concepts. The lasting contribution of Lewin thus may be a way of thinking about problems rather than a theoretical system. This is suggested, too, by the fact that Lewinian students have gone off in many different directions with some identification with a point of view and less adherence to the development of a common set of formal concepts. An interesting discussion in the Deutsch chapter centers about the problem of psychological ecology—namely, how field theory can take account of objective environmental forces, for in field theory these forces must be transformed into the psychological environment. The author accepts the concept of Wright and Barker of the *behavior setting* as "any part of the nonpsychological milieu that is generally perceived by the people of a community as appropriate for particular kinds of behavior." Such behavior settings are coercive because of the shared frames of reference or collective intentions concerning appropriate behavior. This is one way of defining social norms and of accounting for the consistency of the resulting behavior. But it does not help on the problem of ecology. It still must be established that each ndividual subject has the same expectancies as his fellow and that he knows that his fellows have similar frames.

The Hall-Lindzey treatment of Freudian theory and its influence on social psychology is again an excelent summary and interpretation. If anything, it is on the conservative side in evaluating Freudian influence. Even where social psychologists have not accepted Freudian theory as a system they have utilized extensively Freudian concepts and Freudian mechanisms. Though their use of projection, displacement, and identiication have not always accorded with Freudian theory, they have come o fall back upon Freudian dynamics or purposes of explanation. The arly impact of Freudian theory, as n F. H. Allport's *Social Psychology* and E. D. Martin's *Crowd Behavior*, s also minimized in this chapter. It might be added that the tremendous vitality of Freudian theory is due in part to the fact that its mechanisms ave some basis in the actual observaion of human behavior.

In reporting on major trends in ognitive theory, M. Scheerer skillully analyzes some of the basic isues about which stimulus response nd organizational theorists have een in disagreement. The similariies of the concepts employed by ognitive theorists are interestingly developed and are related to many problem areas. This chapter is an important contribution but is perhaps of more value for general psychology than for social psychology. W. W. Lambert had the even more difficult task of relating stimulus-response theory to social psychology. His chapter is better on the presentation of stimulus-response theory than in dealing with the relationship to social psychology. The reasons for the decline of behavioristic theory in social psychology are not faced. One reason for the decline lies in our inability to define clearly the stimulus in social situations. Until we can do this, reinforcement theory will be able to give us explanations only after the fact. Social psychology desperately needs a new objective approach, but Lambert's chapter is an indication that it can not come from the old stimulus-response psychology with its atomism and its grossly oversimplified models.

The most difficult assignment in the theory section was that given to T. R. Sarbin on role theory. While this is a key area for social psychology in relating psychological processes to social structure, there is a dearth of good material on the subject. Sarbin is rightfully concerned with the relation between role and self, but he gives little attention to the relationships existing in a system of roles. It is not very fruitful to consider roles in isolation, for the virtue of the concept is that it takes us beyond the individual into the social system which gives the role its meaning. The concluding statement of the chapter postulates variation in role enactment as a function of at least three variables: validity of role perception, skill in role enactment, and the current organization of the self. This expresses well the orientation of the author and his preoccupation with the peripheral problems of role theory. Nothing is said about the demand character of the role, the sanctions employed in the role system, the function which the role has for the whole system, the competing nature of other roles, the nature of the collective structure of which the role is a part, etc.

The section on research methods is not a complete coverage of methodological and measurement problems, but most of the major areas are included and the treatment represents a high level of competence. The introduction by Allen Edwards on "Experiments: Their Planning and Execution" gives the proper anchorage in experimental design and procedures. F. Mosteller and R. R. Bush follow with an account of selected quantitative techniques which focuses on the needs of social psychologists and which emphasizes nonparametric methods. Just as we require all students to take courses in statistics, so too should we require writers of statistical texts for these prescribed courses to read the chapter by Mosteller and Bush to see how quantitative techniques can be effectively communicated. The chapter on attitude measurement by B. F. Green is a valuable addition to this section and summarizes the assumptions, procedures, and uses of the methods and models in this field. As in the preceding chapter the author is down to earth in his common-sense judgments about the developments in attitude measurement. The two chapters on data collection "Systematic Observational Techniques" by R. W. Heyns and R. Lippitt and the "Interview: A Tool of Social Science" by E. Maccoby and N. Maccoby) are nice combinations of theoretcial analyses and practical considerations of the researcher's problems. The chapter by the Maccobys is noteworthy in cutting through the verbose controversies about techniques and in dealing with the essential dimensions of the interviewing problem. It is valuable too for its reporting of methodological research. The chapter on observational techniques will be very useful because of its inclusion of many different types of observational scales. As the authors suggest, we need to have more investigators utilizing the same observational system if we are to build knowledge in our field. The growing popularity of the sociometric technique and its applicability to social interaction and social structure are recognized in a special chapter on "Sociometric Measurement" by G. Lindzey and E. F. Borgatta. The centrality of the technique to social psychology is emphasized and developments both on the data collection side and on data analysis are described. The authors felt the necessity of choosing between a concept of sociometric technique which followed closely Moreno's early work and a broad concept of the term equivalent to the measurement of social behavior. In choosing the former, they neglected the alternative of accepting the basic technique of reciprocal preferential naming or ranking in the Moreno tradition, but of applying it to a much wider range of problems than personal attractiveness, leadership, and informal patterns in small groups. It is of interest that as this chapter was being written there was already under way a research program carrying the applications of the sociometric technique into new areas, namely, the identification of organizational structure and the processes of communication and control within this structure (the studies of E. Jacobson, S. Seashore, and R. Weiss).

B. Berelson's book on *Content Analysis in Communication* is neatly condensed by its author to describe the procedures, uses, and problems of qualitative analysis. The weakest chapter in the methods section is the one on cross-cultural method. There is little consideration of the methodological problems of dealing with secondary rather than primary data and little attention to the possibilities of controls in the collection of primary data. In all fairness, however, this chapter is representative of the state of the field. Fortunately, the future is more promising in that cross-national research has led to methodological thinking and there is now in preparation a Unesco-sponsored report by S. Rokkan who, with H. C. J. Duyker, has already outlined some of the dimensions of the problem in the *Journal of Social Issues*, Vol. 10, No. 4, 1954.

The final chapter in the research methods section by D. O. Hebb and W. R. Thompson is on the "Social Significance of Animal Studies." Paradoxically, it has as much, if not more, to offer to the social psychologist as any of the chapters in the two volumes in its wise counsel, in its perspective, and in its insights. It affords the stimulation one would expect from the chapter on "Culture and Behavior." The social psychologist can study with profit the development of different order concepts for the description of animal behavior—descriptive categories which grow out of intensive preoccupation with an understanding of the phenomena being investigated. In brief, this chapter on animal studies is a strong argument for the inclusion of a biologically oriented approach to the problems of social psychology.

The second volume deals with social interaction and applied social psychology. Chapter 18, "Socialization" by I. L. Child, and Chapter 26, "National Character" by Alex Inkeles and D. J. Levinson, are particularly impressive. Both of these accounts are thoughtful in their organization and interpretation and resourceful in their suggestions for future research. Child, without sacrificing rigorous standards, faces fully the contradictory findings and the weakness of the research methods in some of the studies, and still makes maximum sense of his material. Research on the socialization process is in a better position for advances because of this chapter. Inkeles and Levinson, in their brilliant account, develop fully the implications of national character both as an independent and as a dependent variable in social psychology. This is a strong chapter methodologically in its specifications for common, standard empirical categories of observation, and in its promising framework for future research. Moreover, it comes closer than any other chapter in adequacy of treatment of social structure variables.

One of the classic chapters in the original *Handbook of Social Psychology*, edited by C. Murchison, was Dashiell's summary of the experimental studies of group effects. This topic is now covered by two chapters which ably represent the new conceptualization of group variables and the new types of group experiments—Chapter 21, "Experimental Studies of Group Problem Solving and Process" by H. H. Kelley and J. W. Thibaut, and Chapter 22, "Psychological Aspects of Social Structure" by H. W. Riecken and G. C. Homans. The work of Kelley and Thibaut is a substantial contribution and, although it neglects some historical origins, is the most comprehensive and understanding account of the experimental literature available. The piece by Riecken and Homans is limited but ingenious. It is an interesting attempt to reduce group activity to a few simple principles, most of which are generalizations from individual psychology. For example, the authors offer the hypotheses: (*a*) a member of a group chooses or likes another member to the degree to which the latter's activities realize the former's norms and values, (*b*) the more one member likes another the more he will originate interaction with him, (*c*) the liking of one member for another is the product of the amount of interaction and the importance of the second member's activities in terms of some group norm. Basic to much of the thinking in this chapter is the old doctrine of self-consistency. Granted the validity of the concept, many social psychologists may want to know how much can be expected from this concept to help in an understanding of group process and group structure.

In keeping with sound tradition there is the usual anthropological piece on culture and behavior. This time it is presented on a grander scale than usual, albeit in somewhat verbose fashion. It is essential that social psychologists should be constantly reminded of how culture-bound their assumptions, concepts, and methods can be. At some point, however, there should be hope o systematic substantive findings in this area as well as the expression of a point of view. C. Kluckhohn has taken a forward step in not only admitting the possibility of universal behaviors but of urging their study. The theoretical position of the article is difficult for the psychologist to follow. On the one hand we are told that anthropologists and psychologists start from data of the same order, namely, human behavior, and that they may even start from the same particular data. The differences come at the level of abstraction and conceptualization. This is a clear and accurate statement. But then we are told that we should consistently and explicitly recognize the interdependence of cultural and psychological phenomena. Here is the source of the real confusion. If anthropology is on a different level of interpretation than psychology of the same phenomena, then we can not talk about the interdependence of the phenomena. If one observer conceptualizes the flight of soldiers in battle as the ascendance of fear motivation and another sees it as the decline of an empire, we can not put these interpretations into the same system of data as interdependent variables. This confusion between the level of data and the level of concepts limits the chapter in its attempts at integrating culture and behavior.

The section on applied social psychology covers four topics: prejudice and ethnic relations, effects of mass media, industrial social psychology, and the psychology of voting. These chapters maintain the high level of reporting and organization of the *Handbook*. It is most helpful to have an article on industrial psychology which sees the problems as field problems in social psychology, as does M. Haire's sophisticated and stimulating presentation. Haire rightly fingers social organization as the focal point in industrial psychology, but he does not take full advantages of the studies and thinking of the workers in this field (Barnard, Simon, Whyte, Worthy, etc.). The Hovland chapter on mass media is a welcome addition to social psychology. It brings together material which otherwise might escape the social psychologist. It is well organized and research oriented in approach so that the interested reader is stimulated to do experiments to fill the indicated gaps.

There are two major omissions in the comprehensive list of chapters, and even these may be considered a reflection of the present state of our discipline rather than an error of editorial planning. The first major lack is the absence of any systematic treatment of social-structure variables. That small face-to-face groups affect behavior is recognized and at the other extreme there is a chapter on culture and behavior. But between these extremes there is no account of the major forces affecting the lives of people save as they are peripherally treated in chapters devoted to some other topic. The most elementary fact about any society is that behavioral relationships are structured and patterned in the form of social organizations, institutions, and groups. Power is exercised as a function of existing social systems, and people live, struggle, fail, and prosper in relation to the social structure of which they are a part. Social behavior does not occur in a vacuum but even the chapter on role behavior treats social roles as if they existed apart from the arena of group struggle and group functioning. The chapter which might be expected to deal with this material, "Psychological Aspects of Social Structure," is concerned only with small groups, and its attempts at general principles take no account of the nature of social structure. Social class, stratification, the exercise of power, the analysis of institutions, the mechanisms of social control, social movements, and change are neglected. This neglect is in part due to the heavy phenomenological bias of present-day social psychology and, correspondingly, of the authors.

The second major omission is the failure to include a consideration of the relationship between personality variables and social variables. Many of the pages are devoted to reducing social problems to personality dimensions. Many group effects, for example, are reduced to the individual's striving for self-consistency. The fit between personality type and social structure is described in the chapter on national character, but other aspects of the interaction effects of personality and social variables are not explored. In other words, we have perpetuated the old dichotomy of approaches: either all individuals are affected similarly by group conditions or all group effects are explained as the expression of personality mechanisms. If social psychology has any unique subject matter, it may well lie in this neglected area of the interaction effects of personality and social settings.

The publication of a handbook makes appropriate some observations about the development of a field in more global terms than is ordinarily the case. When the handbook represents fairly accurately, as does the present one, the principal areas of effort and the major theoretical approaches, this assessment is greatly facilitated. If one were to compare present-day social psychology with the state of the field when the Murchison handbook appeared in 1935, many changes would be obvious. In the first place there has been a marked shift from a behavioristic approach to field theory and to Freudian theory. Perception and cognition have become central topics for many investigators, and where there is an attempt to go beyond phenomenology, the recourse is to Freudian mechanisms and dynamics. In the second place there has been a growth of techniques and methods which are appropriate to social-psychological problems. There has been a corresponding increase in the quantity and quality of research and experimentation. Moreover, social-psychological research has on numerous occasions become a collective enterprise with many large-scale projects utilizing the cooperative efforts of numerous workers. In fact, some of the chapters in this handbook have involved a group of colleagues in their preparation. Then too, many new subject matter areas have been investigated, including political and economic behavior.

Somewhat less obvious among the changes is the origin of research and experimentation. The older social psychology started from everyday problems and attempted to account for the variance of behavior in these situations; experiments were more the miniature of the practical problem, as in the work on prestige suggestion or the accuracy of jury judgments. Today these practical interests continue, but in addition the springboard for research is more often an hypothesis based upon some theoretical insight or attempt at derivation from a theoretical model. This is undoubtedly a healthy sign in the growth of a science. But in the effort to avoid framing a problem in applied terms, it is not necessary to divorce theoretical formulation from the data of our field. The chapter by Hebb and Thompson suggests the importance of intensive concentration upon the phenomena under study. The current trend away from rank empiricism is sound, but it can lead to an overemphasis upon exercises in formal logic at the expense of discovery of significant relationships in social phenomena. Another cause for concern in present trends is the lack of thorough and sustained exploration of a related set of problems so that we can accumulate knowledge in an area and be in a position to make definitive statements about it. Thus far too much effort has been expended at defining or redefining a variable and demonstrating that it is operative at something better than a chance level. Finally, what is still needed is a theoretical system which could handle the problems of social relationships and structure, not as they happen to be reflected in the psychological field of the individual, but as systems of events in the social field.

Received February 1, 1955

HOCH, PAUL H., & ZUBIN, JOSEPH. (Eds.) *Depression.* New York: Grune and Stratton, 1954. Pp. ix+277. $5.50.

Perhaps the main value of this book is that it brings together a wide assortment of material related to depressions and allied conditions. It consists of 16 independent papers originally presented at the 1952 meeting of the American Psychopathological Society. Represented in the symposium are theories and findings from the fields of clinical psychiatry, psychodynamics, anthropology, endocrinology, biochemistry, biometrics, and hospital administration. Attention is given to depressive reactions in children, the aged, cancer patients, and soldiers exposed to isolated Arctic conditions. Most of the articles are reviews of previously published work. As might be expected from a symposium of this nature the individual chapters vary markedly in quality. Kallmann's analysis of genetics in manic-depressive psychosis, Zubin's review of biometric methods in psychopathology, and Steinbrook's cross-cultural evaluation of depressive reactions are outstanding. Some useful criteria for the differential diagnosis of manic-depressive patients are contained in the chapter by Lewis and Piotrowski. Kalinowsky's enthusiastic defense of electric convulsive therapy conveniently ignores the well-established finding—mentioned in an earlier paper—that the present irrational methods of shock therapy, at best, simply hasten the improvement of those who would improve anyway. It may be true, as maintained by Spitz, that institutionalized infants deprived for several months of emotional supplies normally provided by the mother undergo progressive developmental decline, but the absence of adequate experimental controls leaves room for doubt. Might it not be that infants kept in institutions for long periods are biologically less favored than infants who are adopted or placed in foster homes after a brief stay in an institution?

Much still remains to be learned about depressions but in focusing attention on and in providing a wide-range view of the problem the APS has performed a useful service.

JAMES D. PAGE

Temple University

SAVAGE, LEONARD J. *The foundations of statistics.* New York: Wiley, 1954. Pp. xv+294. $6.00.

Only the title of this provocative book will be easy reading for most psychologists. By "foundations of statistics" the author mainly refers to the axiomatic and logical bases of probability theory. After a brief introductiory chapter which notes three major points of view about the interpretation of probability (objectivistic, personalistic, and necessary), a personalistic or subjectivistic theory of decision in the face of uncertainty is developed in formal manner involving an assortment of definitions, postulates, theorems, lemmas, and corollaries. The theory is shown to imply the construct of utility in economic behavior.

The second half of the book is devoted to the impact of personalistic probability on minimax theory, the theory of games, point estimation, interval estimation, and the testing of statistical hypotheses. Overconscientious appliers of common procedures in statistical inference may be somewhat traumatized by the large component of arbitrariness, vagueness, and crudity attributed to the rationale of these procedures.

LEONARD S. KOGAN
Institute of Welfare Research,
Community Service Society of New York

ROBACK, A. A. *Destiny and motivation in language.* Cambridge, Mass.: Sci-Art, 1954. Pp. 474. $10.00.

This is a book of many ideas about language. It will be read with more profit by the layman than by the scholar or scientist because it is written by an amateur (using this word in its exact sense) rather than by an expert philologist or linguist.

Dr. Roback has been interested in language since the age of five when he asked his teacher whether the P'rizzim were the same as Parisians. Had it not been for rebuffs from his language professors at McGill and from Münsterberg at Harvard, Roback might have become a philologist or a researcher in the field of psycholinguistics. In writing this book the author has returned to an early intellectual passion. Appropriately, the book bears the imprint of adolescent exhibitionism and enthusiasm for ideas more than it does the mark of mature and discriminating scholarship. I found it delightful reading.

Destiny and Motivation in Language is a collection of 22 essays on a wide variety of subjects relating to language. I like particularly the chapters on destiny in names, shmoo and shmo, and the psychology of slang. Brutus was destined to be a brute, Boccaccio to be foul mouthed, Kierkegaard to be a gloomy graveyard existentialist, Hitler (Schicklgruber) to be a grave sender, Sir Russell Brain to be an authority on the brain, and the writer of this review to be bald since Calvin means bald. (He is bald!) This is the sort of whimsical, tongue-in-cheek psychologizing that people engage in after a few drinks and rarely take seriously. Perhaps they should!

"Shmoo and Shmo" is a charming essay on the linguistic connection between the name given by Al Capp to his fabulous little cornucopian creatures and the feminine form (shmooe) of the Yiddish word for the erect penis (shmock). Shmock is a hard sounding word; shmooe, shmoo, and shmo are soft sounding. This in essence is Roback's vocosensory theory of the origin of language. The muscular pattern made by the mouth plus the sensory attributes of saying the word contain the meaning of the word. Words sound and feel like the things they refer to. Shmock sounds and feels like an erect penis, shmoo like a soft comforting vagina. Al Capp (born Caplin) unconsciously selected the feminine form of a common Yiddish word, which he must have heard as a child, as the name for his supernurturant birds This essay reminds me of the perspicacious writings of Kenneth Burke on the relation of letter and word sounds to meaning. (How many psychologists are familiar with Kenneth Burke's game with words which he calls "joycing" after the greatest modern wordmaster of them all?)

In my opinion, the most valuable chapter in Roback's book is the one on slang. In this essay, much of which was written forty years ago, Roback manages to sustain a line of thought long enough to reach some pretty interesting conclusions. One such conclusion is that slang is the rebel's form of expression, whether the rebel be an adolescent, a criminal, or a member of a persecuted minority.

Although this book is not an important contribution to the psychology of language, it will be read with interest by people who like the stimulus of fresh, off-beat speculation. Dr. Roback has a keen eye for the ironical, an affectionate feeling for words, and a highly permissive attitude toward ideas. It is pretty apparent that *Destiny and Motivation in Language* was undertaken as a labor of love. It deserves to be read and criticized with the same emotion.

CALVIN S. HALL
Western Reserve University

COOPER, LINN F., & ERICKSON, MILTON H. *Time distortion in hypnosis.* Baltimore: Williams & Wilkins, 1954. Pp. ix+191. $4.00.

In this provocative little book the authors present the results of various investigations of hypnotically induced distortions of time perception and their clinical applications. Those of the readers who have not had the opportunity to read Cooper's earlier publications on this topic will find this book to be, for the most part, pleasant and fascinating reading. For those who are familiar with the aforementioned papers this monograph will be somewhat disappointing since the greater portion of its bulk duplicates the earlier reports. Nevertheless, a fair amount of previously unpublished experimental data has been added and the clinical material is entirely new. In addition, the material dealing with basic concepts and methodology represents an important expansion of the earlier work, and this should prove valuable to experimenters interested in this area.

Essentially, the subject matter deals with slowing down a person's perceived or subjective time through the use of hypnotic suggestions and the effects upon various mental functions. Data are presented which indicate that some functions may be speeded up. It is to be regretted that Cooper and Erickson have preferred to cover a relatively wide range of functions in an exploratory manner rather than limit themselves to a lesser number of more definitive studies. The consequence has been that, although the reported data are superficially convincing, closer inspection tends to leave us wishing for something more substantial. There is a serious lack of objective measures and quantified data, and many of the reported experiments tend to be methodologically oversimplified. The book would also have been more satisfactory if the authors had made a greater attempt to relate their work to current psychology. On the other hand, it should be kept in mind that the topic is admittedly a difficult one to deal with experimentally, and that Cooper and Erickson are essentially pioneering in a new area of investigation. Taken as a totality the reported data do not permit us to categorically ignore hypnotic time distortion and further investigation is clearly warranted.

Cooper devotes some space to theorizing about a general mechanism for the action of hypnotic suggestions. In a somewhat laborious and circuitous manner he reaches the conclusion that hypnosis imbues words with a particular effectiveness to evoke specified experiences via their associated "meaning-tone." This is also the mechanism responsible for time distortion. In this reviewer's opinion, this conclusion hardly seems to elucidate the main question and leaves many other questions unanswered.

In a sense the clinical material reported by Erickson may be the most convincing evidence for the reality of time distortion. The use of the latter seems to be potentially promising for brief psychotherapy and for the treatment of obdurate cases. It also brings to the fore certain temporal aspects of the process of therapy which have been largely ignored to date.

In final evaluation, the reviewer believes that, in spite of certain shortcomings and its rather specialized nature, this monograph will be of interest to a wide group of readers. It represents an original attack upon the problem of temporal perception which cuts across major areas of psychology. The experimenter will discover in it both a new tool and a rich field for further research. Theoreticians, particularly those with a philosophical turn of mind will find food for thought and for speculations. Finally, it gives the clinician and therapist an intriguing new approach to the problems of psychotherapy.

André M. Weitzenhoffer
University of Michigan

McKeon, Richard, et al. *Interrelations of cultures.* New York: Columbia Univer. Press (Unesco), 1953. Pp. 387. $2.50.

This little book deals with eight cultures, particularly their historical origins and value systems. "Interrelations" are treated mainly in historical terms except in McKeon's introductory essay and the final brief statement by a Unesco committee. The authors are predominantly historians and philosophers, and these points of view are central in the discussion. Indian "psychology" is given a few pages (pp. 134 ff.), but they will not appeal to most academic psychologists in the United States. Griaule makes some very suggestive remarks about cognition and learning among African Negroes. There are some excellent—and to this reviewer quite fresh and new—treatments of Indian, Chinese, and African Negro esthetics. Anthropology enters principally into the consideration of African cultures; it gets short shrift on Japan (p. 121).

As always with volumes of this sort, the contributions vary greatly in interest and quality. The two pieces on American culture are extremely thin. On the other hand, Zea on Spanish-American culture is profound and provocative. Atreya on India gives a more compact and intelligible account of central themes than the reviewer has ever previously encountered. Chatterji's characterization of Indian culture also has force and point, though one feels he strains to present it in a manner that will fit best with certain currents in Western thought.

In general, this volume will attract only those psychologists who are interested in the cross-cultural study of values or those who enjoy somewhat discursive essays seasoned with occasionally exotic facts. Yet the best of these chapters are extraordinarily thoughtful syntheses of vast bodies of material, and Unesco deserves thanks for making them available.

Clyde Kluckhohn
Center for Advanced Study in the Behavioral Sciences

Ziskind, Eugene. *Psychophysiologic medicine.* Philadelphia: Lea and Febiger, 1954. Pp. 370. $7.00.

Here is a book written for the general medical practitioner that will interest psychologists more by its purpose than by its content. Its premise is that one-third of the patients seeking medical aid have symptoms solely or primarily due to psychogenic pathology and another third combine psychogenic with physiogenic pathology. The requirement for psychodiagnosis and psychotherapy in the treatment of general medical disorders is clear indeed.

Dr. Ziskind, head of the department of psychiatry and neurology at the Cedars of Lebanon Hospital in Los Angeles, has been coordinating a training program for physicians to meet this need. He drew heavily on this experience in preparing the portions of this book dealing with diagnosis and therapy.

The book consists of four parts. The first one provides background and orientation to the problem. The second part is concerned with diagnostic and psychotherapeutic technique. The third part consists of a review of familial and other social factors in psychopathology, plus a general treatment of psychiatric syndromes and common psychiatric emergencies with which the physician may be confronted. The fourth part includes a useful review of contemporary schools of psychiatric thought.

The writing style is easy and lucid. In keeping with the aim of the book, there is an almost "do it yourself" quality to it. This is one of its shortcomings. Skill, personality, and character requirements for effective psychotherapy are stated as precepts, as if the reading by the physician is a sufficient fulfillment of the requirements. The requirements themselves are not at issue. Ziskind states that the personality and attitude of the therapist must be encouraging. He should be kindly, sympathetic and interested, permissively nonmoralistic, nonjudgmental, and nonauthoritarian (p. 149).

Within the framework of a supervised training program at a hospital there is some modest assurance that these requirements may be met. As a minimum, a physician lacking these characteristics can be discouraged from further participation. Ziskind reports that the hospital program has been highly successful, and it may be assumed that the success is due in no small measure to the availability of a well-trained psychiatric staff.

For a training program of this type this book would be a useful systematic guide. However, the book is addressed to the practitioner at large. The wisdom of its use independently of supervision will be seriously questioned. Ziskind believes that because psychophysiologic illness is so rampant, the physician is obliged to deal with it as best he can anyway and after reading this book the physician can approach this task with less anxiety.

What is at issue is whether the patient is better off if he is exposed to the risk of therapeutic mismanagement. This is the dilemma posed by Ziskind's effort. The merits of his effort are so obvious that it would be highly valuable to assess the magnitude of the risk empirically even though this may be a large undertaking.

Against such data as a background, other solutions may also be considered. These would include the overdue modification of medical school curricula and the establishment of a more cooperative working relationship between the medical and psychological professions.

JOSEPH E. BARMACK
College of the City of New York

MURPHY, G. & BACHRACH, ARTHUR J. *An outline of abnormal psychology.* (Rev. Ed.) New York: Random House, 1954. Pp. ix+597. $1.45.

This is a revision of Dr. Murphy's 1928 contribution to The Modern Library Series and almost everything but the title has been changed. Dr. Murphy gives most of the credit for the selection of the material to his coeditor. The first chapter is a Gardner Murphy original. In addition to chapters on the dynamics of schizophrenia, psychoneuroses, and paranoid disorders, there are chapters on specific topics such as adolescent conflicts and adjustment, obesity, the addictive drinker, psychological factors in epilepsy, the individual and society, Rorschach variables and ESP scores, mongolism, and the treatment of the withdrawn child. The editors have selected published material which has breadth, reader importance, technical accuracy, and simplicity of report. The unity of the book is striking, in view of the wide range of subject matter and the sources of the papers included. An appropriate and helpful glossary is appended. The volume would have been improved by an index, especially since it is suitable for college classes.

This is not just another book on abnormal psychology, as the title may suggest, but a collection of readings which may be used either to supplement existing texts, or alone, as an aid to stimulating insight and understanding in the searching layman. As psychology's role in the community continues to be enlarged, there will be more and more need for authentic, readable, and relevant books such as this for the average reader.

FRED MCKINNEY

University of Missouri

HOIJER, HARRY. *Language in culture.* Chicago: Univer. of Chicago Press, 1954. Pp. xi+286. $4.50.

The late Benjamin Whorf's dramatic presentation has interested a large part of the intellectual world in the hypothesis of "linguistic relativity." Whorf maintained that the cognitive style of a people is in correlation with, and perhaps determined by, the lexicon and structure of the language spoken by that people. In 1953, at the University of Chicago, a group of scholars representing linguistics, anthropology, psychology, and philosophy discussed for one week the "relativity" thesis of Whorf. The report of the conference comprises seven papers and nine discussion sessions.

With Whorf, as with Freud, no reading, however careful, can tell us exactly what the author meant if exact meaning implies operational definition. The exegesis of Whorf is a principal topic of the conference. This portion of the discussion may seem to be wasted since the problem of Whorf's exact intentions is, after all, a rather trivial historical question. In fact, however, each discussant takes this opportunity to propound his own views and receive criticism of them, letting Whorf absorb the odium for whatever may be said that cannot survive criticism.

Whichever of several things Whorf meant to say the conference members generally agree that his data (largely from his study of Hopi) are suggestive rather than conclusive. "Linguistic relativity" is a collection of hypotheses about language, culture, and cognition. The conference accordingly gives some time to the design of research intended to test these hypotheses. Hoijer makes an interesting proposal for comparing general cultural similarities with linguistic affiliations among the Navaho, Hopi, and other American Indians. Hockett gives the best feeling for method and also provides valuable examples in his comparison of certain limited features of English and Chinese. In a final discussion research ideas are rather carelessly tossed out. Several of these are promising (notably those of Lounsbury and Kaplan) but the reader is disappointed by the failure of the conference to think through any one idea to a point where its practicability and merit can be judged.

Hoijer has condensed and edited

some 919 pages of typescript to obtain the 150 pages of discussion that appear in the book. These discussions read smoothly and are generally intelligible. However, the transcribed diaglogue of ideas never has the sustained direction and thoughtfulness of a literary construction. The unfortunate fact is that not even every other idea expressed in a conference is worth recording. As a result this book makes an extremely uneconomical presentation of a small number of good ideas.

ROGER W. BROWN
Harvard University

KLOPFER, B., AINSWORTH, M. D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique.* Vol. I. *Technique and theory.* Yonkers, New York: World Book, 1954. Pp. x+726. $6.50.

There is something for everyone in this first volume of a two-volume work on the Rorschach. Probably the primary use of the book will be as a teaching manual for students learning the mechanics of scoring and the feeling for interpretation, but even experienced users of the technique may read with profit the sections on theory and interpretation.

The sections on administration and scoring are written clearly and numerous examples are provided to clarify scoring problems. Few changes in the scoring system presented twelve years ago in *The Rorschach Technique* have found their way into the present procedure. There is some clarification of methods of scoring shading responses though these distinctions are likely to continue to be frustrating to students. Ten popular responses are listed—the same ones as twelve years ago, and a slightly different list from that offered by other experts. The form-level rating procedure, an attempt to give a numerical score for each percept reflecting its accuracy, will continue to baffle statisticians and others with a bias against adding, subtracting, and averaging cabbages and kings.

The sections on interpretation are well written and will stir little controversy among experienced users of the technique. Some readers are certain, however, to criticize the failure of the authors to support their statements with data in the literature. The authors, anticipating this objection, present their interpretations as "hypotheses" and indicate that the usefulness of these hypotheses has been established in clinical practice and that they may be used "pending the conclusion of the extensive investigations that will be necessary to evaluate their relative validity." Apparently existing validation studies do not fit their criteria. Certainly clinical use has resulted in few changes in interpretation. The authors also indicate, wisely, that "the integration of the findings into some sort of meaningful dynamic picture depends largely upon the examiner's basic understanding of human personality." Again, there is much in the literature which could be used to document this hypothesis.

The section on validation problems makes a case for the point of view that "there is no sharp dividing line between validation research and the clinical use of the Rorschach technique . . . indeed, the technique itself is continually changing through the validation process." A test of this latter hypothesis through a comparison of scoring and interpretation procedures by the same authors over the years, and by an examination of possible gradual rapprochement between different schools, would possibly raise some doubts here. The section on validation concludes with the statement that "validation research should be planned and the results discussed by those who have more than a superficial understanding of the basic concepts involved in Rorschach interpretative hypothesis. . . . " Few will disagree, though the statement leaves a nice opening for a rejoinder by the statisticians!

To single out these statements does not give an accurate picture of the generally excellent discussion of the problems of validation and of existing approaches to these problems.

Also included is a brief, though stimulating chapter on implications of contemporary personality theories, a somewhat more esoteric chapter on ego psychology, a long case study, and a Rorschach Prognostic Rating Scale (experimental) designed to predict a patient's response to psychotherapy. This latter *may* be just what Eysenck has been looking for, though again it may not. It would be interesting to hear him discuss it with the authors at a symposium. At the end of the book, as an appendix, is an added chapter on Freudian and Jungian approaches to the development of ego and self. The author is Jack Fox, who is not represented on the masthead, perhaps because there is no mention of the Rorschach in the chapter.

The book contains no general list of references though there is a bibliography at the end of many chapters. Those interested in or favoring particular cards of the Rorschach will find a list of the pages on which each card is mentioned.

In a work of this scope it is easy to find statements or points of view with which one can disagree. Also with a second volume in the offing one may find his objections concerning lack of some form of content answered with the companion volume. There is little doubt that the present work will find a wide audience and deservingly so.

GEORGE W. ALBEE

Western Reserve University

KATZ, D., CARTWRIGHT, D., ELDERSVELD, S., & LEE A. McC. (Eds.) *Public opinion and propaganda.* New York: Dryden, 1954. Pp. xx+779. $6.25.

The Society for the Psychological Study of Social Issues has produced an excellent book of readings on public opinion and propaganda. In keeping with its interdisciplinary orientation, a wide range of the relevant literature in political science, history, anthropology, sociology, economics, and psychology has been culled to provide a selection of readings (74 in all) showing the possibilities of the empirical approach to problems in the area as well as the societal context, the political structure, and the social-psychological dynamics of opinion formation. There is as wide a coverage of points of view, types of studies, and problems, both practical and theoretical, as one has a right to anticipate in a book of readings intended for courses in public opinion and propaganda. On the basis of the sampled advanced work included in the collection, one wonders, however, whether the general area of public opinion and propaganda will acquire that eventual "scientific" status which the editors suggest such work may adumbrate. "Control" and "number" are not enough. After all, Ptolemaic theory also had its systems of quantification.

IVAN D. LONDON

Brooklyn College

McMILLAN, B., GRANT, D. A., FITTS, P. M., FRICK, F. C., McCULLOCH, W. S., MILLER, G. A., & BROSIN, H. W. *Current trends in information theory.* Pittsburgh: Univer. of Pittsburgh Press, 1953. Pp. xiii+188. $4.00.

This is the seventh in a series of annual conferences on *Current Trends in Psychology* arranged by the Department of Psychology of the University of Pittsburgh. The seven participants and the titles of their contributions are: McMillan, "Mathematical Aspects of Information Theory"; Grant, "Information Theory and the Discrimination of Sequences in Stimulus Events"; Fitts, "The Influence of Response Coding on Performance in Motor Tasks"; Frick, "Some Perceptual Problems from the Point of View of Information Theory"; McCulloch, "Information in the Head"; Miller, "Information Theory and the Study of Speech"; and Brosin, "Information Theory and Clinical Medicine (Psychiatry)."

As is frequently the case with such compendia, this one is rather a hodgepodge with different authors trying to do different things in different ways. The unifying theme, such as it is, is modern communication theory. McMillan attempts a brief and general account of communication theory, in which attempt he is very successful. It is gratifying to know that mathematicians *can* communicate with psychologists if they want to.

Grant talks about a number of experiments dealing generally with partial reinforcement and the discrimination of stimulus frequencies occurring with various probabilities. Although the experiments are interesting in their own right, the attempt to relate the results to information theory is strained. It is hard to escape the feeling that it would have been much less cumbersome to talk about these studies in other terms.

Fitts is concerned with the speed and accuracy of discrete and continuous movements and with stimulus-response compatibility. He reports some exceedingly ingenious experiments but it is difficult to know how much their conception is in debt to information theory. At any rate, it is possible to argue that the principal contribution of information theory to these results has been to suggest the use of logarithms.

McCulloch has an interesting survey of some neurophysiological problems of sensation and perception and Brosin talks about some problems in clinical psychology and psychiatry. In both cases the principal connection with information theory has been the translation of some common psychological and physiological terms into some uncommon ones, e.g., "quantize," "modulate," and the like. If both authors had applied terms ordinarily used in their respective fields, their discussions would undoubtedly have contained as much information and would probably have been better communications.

Of all the authors only Frick and Miller describe some classes of psychological experiments to which information theory has been able to contribute something other than a new vocabulary. Miller is especially levelheaded in his appraisal of the usefulness of information theory for psychology and a paraphrase of the first two pages of his contribution might well summarize the outcome of this conference: There are not yet many places in the study of human behavior where information theory can be profitably applied.

Since the chances are good that you do not agree with the reviewer's appraisal of information theory, you ought to read the book yourself. It is worth reading.

A. CHAPANIS
The Johns Hopkins University

ZILBOORG, G. *The psychology of the criminal act and punishment.* New York: Harcourt, Brace, 1954. Pp. xi+141. $3.50.

This short book grew out of a Yale lectureship under the joint auspices of schools of law and medicine. It examines the motivation and consequences of punishment of the criminal and points to the "curative powers of healthy self-punishment which never becomes hostility."

It will make its major contribution to the practicing professions, but psychologists acquainted with the hypotheses reviewed here will welcome their emphasis by a writer who is well regarded in his own and related fields. Zilboorg presents excellent suggested courtroom ethics for the representatives of the behavior sciences.

FRED MCKINNEY
University of Missouri

WOLSTEIN, BENJAMIN. *Transference: its meaning and function in psychoanalytic therapy.* New York: Grune & Stratton, 1954. Pp. xiii+206. $5.00.

Trained in philosophy, psychology, and Sullivanian analysis, Wolstein has given thought to the rich source of fantastic data brought to light chiefly by methods stemming from Freud: dreams, free association, "transference" observations—areas largely avoided by experimentalists. He also understands operationism and precise models of science. His book demonstrates well the feasibility of bringing into scientific consciousness those analytic concepts which sometimes seem to have come up dripping with the amorphous illogic of their dreamworld origin.

His review of Freud's and later analytic thinking suggests that Freud saw transference as infantile behavior toward parents "transferred" to a present situation. Wolstein sees transference in terms of observable here-and-now behavior. Freud pursued relentlessly the reconstructed memory of an event from early infancy which "determined" the neurosis; using first hypnotism, then strong suggestion, and finally an authoritarian exploitation of "positive transference." Negative transference he battered down as resistance, and countertransference was not admitted to the patient. Wolstein accepts extrapolative reconstruction of infantile events, but also emphasizes later developmental stages. He keeps the focus on present interactions with careful attention to the therapist as an interbehaving organism (countertransference) rather than as a hypnotic authority.

The book reflects the groping trends of opinion developing from therapeutic experience and from increased awareness of the problems of definition. Many questions related to transference are brought closer to the status of answerable questions.

ROY M. HAMLIN
Western Psychiatric Institute
University of Pittsburgh

WEVER, E. G., & LAWRENCE, M. *Physiological acoustics*, Princeton: Princeton Univer. Press, 1954. Pp. xii+454. $10.00.

This new volume from Professor Wever's laboratory can be immediately recommended to two classes of readers: the few who eat and sleep audition, and the many who need a clearly written text to keep them posted on the latest facts plus authoritative comment from a logical mind on the meaning of those facts.

For workers already immersed in the area, the book provides two values. It brings together much of the material Wever and his associates have scattered over scores of papers in the acoustical, otological, and psychological literature; and it also presents new data and fresh comment on old data, both of the very greatest interest.

The plan of the book is to examine the transmission of acoustic energy from air-borne vibrations clear through the system up to but not including the primary neuron. After the three short introductory chapters, well over two-thirds of the remaining 375 pages is devoted to the system up to but not including the cochlea. This emphasis on the drum, the ossicles, the tympanic muscles, problems of bone conduction, of the relative efficiency of routes to the cochlea, and of clinical problems in sound conduction form a salutary change from a common notion that "auditory theory" concerns only basilar mechanics and associated neurology. There is much besides a dramatic concentration on cochlear patterns to engage the attention of an auditory theorist.

This book makes clear how Professor Wever has focused the full attention of his laboratory for a quarter century on the whole problem of sound conduction. Beginning at the eardrum and proceeding to intracochlear electrodes, every phase of the subject has been touched upon by him in some, often crucial, experiment. The result is a treatise in the best tradition. Certain problems unsolved since Helmholtz have yielded to Wever's experimental technique. A good example is his chapter on the eardrum as a conical lever. He first summarizes Helmholtz's analogue of the catenary as the principle by which the force applied to the malleus is amplified, marshals the anatomical evidence on the elasticity of the radial and circular fibers, and then repeats Helmholtz's crucial experiment with a technique enabling him to move the manubrium of the malleus accurate to .0005 mm. As a result, auditory scientists will never again seriously consider the eardrum as an amplifier.

As an example of the care and logic exercised in Wever and Lawrence's work, one may cite the series of studies on the function of the round window. Although Eduard Weber long ago had the generally correct notion, numerous attempts by several modern laboratories to document this notion had proved ambiguous. It remained for Wever and Lawrence to concentrate their attack on the round window with quite nice techniques of blocking, probing, loading with an air column, with and without ossicles present, and to show in precise quantitative fashion the proportionate contributions of mass and stiffness to the system, as they affect high-tone and low-tone responses. Thus not only is the main function of the round window logically proved, but the details are expressed in quantitative fashion and a rational explanation offered. Finally, the failure of earlier studies to produce conclusive results is very civilly given by inference, and a true explanation for their data is at the reader's hand for the first time.

The medical significance of Wever's decades of work on sound conduction is best appreciated by otologists, to whom he often personally addresssed himself and who have taken the opportunity to underwrite some of his efforts.

To psychologists, the features of Wever's work of most significance are his discovery of the cochlear microphonic (with which of course the name of C. W. Bray is always connected as well), and his insistence on the unsatisfactory nature of "place" as the only principle of cochlear analysis. In his book *Theory of Hearing* he fully discussed the reasoning which led him never to neglect the fact that a nerve fiber always conveys information by its periodicity as well as by its anatomical location, and which led him to assimilate the characteristics of both space and time in a unified theory. The wording here is reminiscent of Harvey Fletcher's space-time theory, but there are real differences. In Wever's view a fast-acting pressure wave sets up a geometric vibration pattern on the basilar membrane, and this pattern is sufficient to serve as the intermediary between the oscillatory pattern of the stapes and the space-time pattern of nerve response. In Fletcher's view, and in that of many investigators today, the pressure wave serves mainly to initiate a second, much slower, traveling wave moving up the basilar membrane much like a vibration along a shaken rope, and the characteristics of this traveling wave determine the neural space-time pattern.

It must be said that direct observation and certain intracochlear multiple-electrode studies seem to support the existence of a relatively slow traveling wave. Nevertheless, Wever and Lawrence offer the possibility of another interpretation, and do not regard the wave as proved beyond question. In addition, they offer a most ingenious and carefully controlled experiment showing that the cochlear pattern is very similar no matter whether sound is introduced by a tube over the oval window or by a tube over a little hole drilled in the apex. It is admittedly most difficult to explain this result by a slow-traveling wave, as indeed it is difficult to explain many of the extremely rapid phenomena of audition with a slow wave of any sort. In these matters, Wever's theory comes off very well.

Many will wonder just how and in what region Wever would have the frequency principle for low tones give way to the place principle for high tones; Wever will simply state that such questions are properly the subject of experiment not argument. Moreover, it may be that the experiment of introducing sound at the apex is not crucial; one can believe, for example, that the construction of the apical end of the basilar membrane and associated spiral ligament is not conducive to being snapped like a rope, and that the pressure wave beginning at the apex and hitting the oval and round windows within about 20 microseconds, is reflected back up the canals and initiates a slower traveling wave at the usual end of the basilar membrane. Thus a sine wave impressed upon the fluid at opposite ends could well initiate two traveling waves following each other with only a slight phase lag, well within the limits of measurement, and it would be possible to create cancellation effects quite generally throughout the canals, as indeed Wever and Lawrence found.

Suffice it here only to say that the data on the question as to the emergence of the pattern of excitation in the cochlea, the very voluminous data from several laboratories since the publication of his book of five years ago, have not caused Wever to change his mind as to the necessity for a synthesis of both principles, space and time. And in the discussions which are now progressing in the several journals, Wever, by his persistence and elegance in the laboratory and in the literature, makes quite as good a case as anyone.

J. DONALD HARRIS
U. S. Naval Medical Research Laboratory,
New London, Connecticut

HALL, CALVIN S. *A primer of Freudian psychology.* Cleveland: World, 1954. Pp. xii+137. $2.50.

Hall's small and unpretentious volume is in most respects just what the author and publisher claim it to be—a relatively nontechnical, yet systematic primer of Freud's theory of personality. A second claim—that its material is selected from and reflects Freud's dynamics of normal rather than abnormal psychology—seems to be a bit overdrawn. Although Freud certainly distinguished between normal and abnormal states, it is equally certain that he did not invoke different dynamic constructs to account for each of them. However, the structure of Hall's book is not vitiated by the artificial dichotomy of dynamics which he introduces, nor is its value lessened as a singularly clear, comprehensive, and unadorned account of essential Freudian theory.

In his introductory chapter Hall gives a brief and informal picture of the steps which led Freud from biological research to psychology. Apart from this account in which his admiration for Freud, the Scientist, is frankly explicit, the author's own views are never obtrusive. They are in fact, hardly discernible. This anonymity of "reporting" is one of the many positive features of the Primer. Hall's careful organization of topics, his succinct introduction to each major new one, and his brief but precise summaries should make this book a prototype.

The lay reader will get a reasonably good overview of Freudian theory from the book and will be assisted further by the list of recommended readings furnished him. For the student, the book should provide an excellent frame of reference for further study. The topical list of references to Freud's writings will be of real assistance to him.

DAN L. ADLER
San Francisco State College

FAIRBAIRN, W. RONALD D. *An object-relations theory of the personality.* New York: Basic Books, 1954. Pp. xi+312. $5.00.

In a series of lectures delivered in the 1940's, Fairbairn developed what for him was a renovation of Freud's libido theory and Abraham's "developmental phase" theory. It is these lectures which appear in the first portion of this book; in succeeding sections there appear three case histories which are used to illustrate the author's resynthesis, and a group of miscellaneous papers.

If one were to use the criterion of internal consistency in evaluating the theoretical contribution of the author, he would come off very well indeed. From lecture to lecture, Fairbairn unfolds a program of interpretation which is essentially free from *non sequitur*, and carefully avoids the loose ends which often characterize this method of presentation. The premise is stated in the observation that schizoid states are found in practically all psychopathological cases, and by implication must develop early in life. They occur, therefore, as resultants of varied relationships between the person and objects which he seeks in response to libidinal prompting. Using these "object-relations" as a key concept, Fairbairn proceeds to reorganize classical interpretations of such things as melancholia, impulse-psychology, and the holy trinity of ego, id, and superego.

If one were to evaluate this New Look by other than consistency criteria, however, it is doubtful that the concensus would be so favorable. For many *classical* psychoanalysts, Fairbairn will probably represent a particularly repugnant type of heresy in his disavowal of "accepted" basic principles. To the neo-Freudian, his dynamics may appear too far removed from the social world to be considered practicable. Finally, for those who demand of a conceptualization that the referents be clear, and that labels and names not be treated as entities, there will be little incentive to read beyond the first chapter.

Among the miscellaneous papers in the last section there is one which might be read as an object lesson in righteous indignation. Fairbairn argues here that there is strong discrimination against the acceptance of psychoanalysis (and surrealism) in academic curricula, from which it follows that "... the higher the cultural value of a subject in the university curriculum, the less freedom of thought and enquiry is permitted" (p. 253). The scale by which he measures cultural values is not described.

DAN L. ADLER
San Francisco State College

SHELDON, WILLIAM H., in collaboration with DUPERTUIS, C. WESLEY, & MCDERMOTT, EUGENE. *Atlas of men.* New York: Harper, 1954. Pp. xvi+357. $10.00.

This Atlas is organized as a working guide to be used in conjunction with Sheldon's earlier volumes, *The Varieties of Human Physique* and *The Varieties of Temperament.* It contains somatotype photographs of 1,175 men, selected as representative of the 88 somatotypes identified in the total sample of some 46,000 men ranging in age from 18 to 65 years. As the book is primarily a supplement to the earlier volumes, it is necessary to turn to them for detailed descriptions of the criteria for ratings, unless one can learn, by repeated leafing through the photographs, to identify the differentiating factors between, say, a 2 and a 3 in ectomorphy. Each somatotype, in addition to photographs of several examples at different ages, is presented with (*a*) a thumbnail sketch, largely in fanciful analogy to an animal "totem" that is intended to characterize both the physique and its personality, (*b*) a chart showing the relative position of this and similar somatotypes, (*c*) a description, including the incidence of this type in the population, and the kinds of pathology, both physical and mental, that are most characteristic of it, and (*d*) a table giving weight for height and age, and a chart of the $H/\sqrt[3]{W}$ index by age, for purposes of evaluating nutritional status according to somatotype.

Appendices include a glossary; a set of tables for distributions of somatotypes, by age, according to the $H/\sqrt[3]{W}$ (which may be used as a guide for identifying somatotypes); a description of equipment and procedures (including a chart for converting weight and height into the $H/\sqrt[3]{W}$ index); and a table giving the case numbers of subjects in the Atlas having varying degrees of "gynandromorphy" and "texture," the two relatively independent variables.

New data on this system of somatotyping are discussed briefly. In a consideration of the sex factor, Sheldon defends his use of the same standards for both sexes, and presents charts showing the differences in distribution of the somatotypes of men and women. The females predominate in the areas of endomorphy-ectomorphy, with no ectomorphic mesomorphs, while the males have a greater variety and spread of types, the mode being high in the mesomorphic component. An Atlas of Women, a companion volume to this one, is in preparation. In dealing with the age factor Sheldon restricts his discussion to the span 18 through 65 years, but we find that an occasional boy under this age, as young as 11 years, has been slipped in among the illustrations. This seems to me a dubious procedure, because body-proportions change so markedly in boys between 11 and 18 or 21 years. Lacking longitudinal data on the men, the author has made use of his subjects' reported weight histories, together with such casual photographs as they could supply, to project the curves backward in time. Each somatotype has a characteristic curve, as some naturally put on more weight than others. These curves seem plausible, but they need verification by actual repeat measures and assessments of the same men as they grow older. There is a discussion of the effect of nutrition, in which Sheldon defends his system against critics who have shown that such things as semistarvation alter the apparent somatotype.

For those who are concerned with the problem of physique and its relations to personality and to disease, this book is provocative and a challenge to further research on the hypotheses and assumptions presented. For those who use the Atlas as a tool or guide for evaluating subjects, diligent study of the many illustrations should contribute to proficiency in somatotyping. However, it would be easy for the uncritical to overvalue the validity of Sheldon's assumptions. Although at the outset he is careful to call his work a "pilot study" and to point out the need for more validating procedures, the descriptions of each somatotype carry the ring of certainty, e.g., of somatotype 1 3 6 he says, "Flycatchers. Frail lightly built birds who yet are fierce-hearted hunters. . . . When found in the mental hospital . . . the 1 3 6 is most often labeled *paranoid schizophrenia*. . . . But if the 1 3 6 is sometimes paranoid he is often heroic. For wherever paranoia dwells, there also is the home of heroic devotion" (page 45).

NANCY BAYLEY
National Institute of Mental Health

HENRY, ANDREW F. & SHORT, JAMES F. JR. *Suicide and homicide.* Glencoe, Ill.: The Free Press, 1954. Pp. 214. $4.00.

Subtitled "Some Economic, Sociological and Psychological Aspects of Aggression," this book is by two sociologists: Henry is from Harvard and Short from Washington State. It treats suicide and homicide as dependent variables and seeks to relate them to the business cycle, forms and degrees of social control, and aggressive tendencies against self or others. In five appendices they appraise previous research, examine statistical techniques for studying business cycles, and use these techniques to study the relations between cycles of suicides and of business, and offenses against property.

Henry and Short conclude that suicide rises in depression and falls in prosperity whereas homicide rises in prosperity and falls during depression. Both these tendencies are associated with social status—suicide with high and homicide with low status. As regards sociological variables, the authors study what they call "strength of the relational system" defined in terms of marital status, urban-rural residence, and ecological distribution. They conclude that the greater the strength of the relational system, the lower the suicide rate and the higher the homicide rate.

In the chapter on psychological determinants, Henry and Short study the differential superego structures of the suicidal and the homicidal and find that suicide is correlative with high guilt and homicide with low guilt.

This ambitious work by Henry and Short aims to build on Durkheim's theory and the empirical research of more recent Americans. The interrelation of the types of variables considered by the authors seeks to solve questions of the association of superego structure with the combined effect of economic stress, social status, and the strength of the relational system. Thus, for example, the authors find that, with status held constant, the relationship between suicide, homicide, and the business cycle indicates that high status categories experience frustration during business contractions while the lowest status category experiences frustration during business expansions. The vast body of material that Henry and Short seek to handle in short compass leaves one with the impression that more has been studied than has yet been assimilated, but that broad avenues of investigation have been opened up. Since in studies of suicide and homicide all associations are best shown in terms of rates, we are forced to reflect on why everybody in certain categories—age, marital status, socioeconomic status, race, ethnicity, religion—where rates have been differentially established does not commit suicide and homicide. We might then be forced to conclude that certain personality types are suicide-prone or homicide-prone and then would have to seek to discover how and why these types show up in different ascriptive categories and what differential socializing patterns are involved in those prone and those not prone.

GEORGE SIMPSON

Brooklyn College

LOWENFELD, MARGARET. *The Lowenfeld Mosaic Test.* London: Newman Neame. 1954. Pp. 349+144 color plates. 50s. (available in U. S. from Psychological Corp., New York City)

The Lowenfeld Mosaic Test has been in use for twenty-five years, and articles about it have appeared for fifteen years, but this is the first full-length treatment of the test. It has been long overdue, since the test is now rather widely used. This book will probably do for the Mosaic Test what the English translation of the "Psychodiagnostik" did for the Rorschach: it will transform it from an esoteric technique into a popular clinical instrument. And, as in the case of the Rorschach, some psychologists will not be too happy about this.

The resemblance of this volume to the "Psychodiagnostik" is marked. Consciously, or unconsciously, Dr. Lowenfeld seems to have been influenced by Rorschach in the approach to the study of personality, the medium employed, the rationale, the intuitive appraisal of personality, the use of single cases as prototypes, and even the important elements of the test, such as color, form, movement, space, content, and many others. An important difference is that in the Rorschach the subject is faced with completed designs while in the Lowenfeld test, the subject constructs the design.

The book is divided into two parts: Part I deals with the description of the test, including the origin, administration, and principles of classification of designs produced by subjects. Part II describes the use of the test with children, subnormals, normals, neurotics, organics, and psychotics, and discusses its use in the study of cultural problems.

Diagnostic procedures are presented tentatively throughout, and the reader is frequently cautioned to use this test experimentally rather than as a finished instrument. In fact, a chapter on scoring procedures, apparently written for this book, was not included because the author does not wish to freeze scoring methods and norms at this point, since new evidences of the effects of cultural factors in the use and interpretation of the test are constantly coming in, particularly from the United States. The latter results seem to differ markedly from those obtained in Europe, where most of the early work was done. In spite of the cautions constantly presented, the impact of the "typical" cases and the large number of case descriptions leaves the impression that the test can be used with a reasonable degree of assurance as a diagnostic clinical instrument.

The chapters dealing with the description of the test convey the impression that the designs produced by subjects can be readily classified into stated categories which lead to almost automatic diagnosis. On the other hand, the parts of the book which concern themselves with the use of the test bring the reader back to reality with a jolt. Warnings about the difficulties in classification experienced with live cases, examples of borderline and ambiguous designs, samples of overlapping diagnostic possibilities for similar designs, instances of subjective interpretation, and conflicting reports by different workers abound in the many case samples presented, and serve to disabuse the reader of any hopes for a reasonably simple and useful technique for personality appraisal, in the event that he had entertained such hopes.

In spite of twenty-five years of experimentation and clinical use, this test is still in its beginning stages. It is now ready for validation. There is no doubt about the fact that competent clinical psychologists will find this a useful projective technique—an important addition to the clinical battery—if they are not already using it. But then, competent clinical psychologists can find almost any series of samples of human behavior useful for diagnostic purposes, particularly one that has had considerable use, and therefore presents some basis for comparison of results. Those who insist on "respectable" methods of objectification, and on standard approaches to reliability and validity will, no doubt, be as violent in opposition to this test as they are toward other projective methods.

The 144 separate plates, beautifully printed in color, constitute, in effect, approximations to norms for the Mosaic test, although this is disclaimed by the author. They add a great deal to the usefulness of the volume.

MORRIS KRUGMAN
New York City Board of Education

BELLAK, LEOPOLD. *The Thematic Apperception Test and the Children's Apperception Test in clinical use.* New York: Grune and Stratton, 1954. Pp. x+282. $6.75.

More than anything else this book gives evidence of the extensivity and maturity of Bellak's clinical experience. Both in the expository sections of the volume, in which the author elucidates his thinking about such items as theoretical foundations for projective testing and the use of the TAT in psychotherapy, as well as in the case illustration sections in which sample TAT and CAT stories are analyzed and interpreted, there is a richness in clinical wisdom and an erudition in psychoanalytic personality theory. However, there are shortcomings of the book as a *manual* for TAT and CAT interpretation, unless one wishes to accept Bellak's interpretive procedure "lock, stock, and barrel."

Bellak's basic premise rests on his statement: "Since I believe that the strength of the TAT lies in its ability to elicit the content and dynamics of interpersonal relationships and the psychodynamic patterns, my method of interpretation and my scoring categories are primarily concerned with these dimensions and only to a small extent with the formal characteristics." Based on this conviction, he presents his method of analysis for the TAT using such factors as "main needs of hero," "nature of anxieties," "main defenses against conflicts," and "integration of the ego." A different, but somewhat similar schema is suggested for the CAT. The examiner is then expected to develop inferences following this type of analysis of thematic stories, based on some general principles that are proposed.

There seems to be little doubt that, according to present knowledge, the evaluation of projective test results *for an individual* will require a clinical interpretation of both qualitative and quantitative aspects of the test protocol *plus* an integration of such derived hypotheses with the life history and/or clinical behavior of the individual, as a minimum. Nevertheless, one may question the value of Bellak's schema on several counts, not necessarily of equal significance: (*a*) There is no necessary relationship between his projective theory and his method of analysis. The bridge is a tenuous one. (*b*) No evidence is presented concerning the scoring reliability for the suggested "scoring categories." Many of these apparently require considerable clinical skill, and may, presumably, be evaluated differently by different examiners. (*c*) One may be fascinated by the nature of the scoring categories but one may also wonder what the relationship of each of them is to the observed or latent personality characteristics of the examinee. Bellak has frequent recourse to the phrase "in my experience," but this type of evidence may not be convincing to other clinicians. (*d*) There are no clear bases for judging the "adequacy" of such factors as "severity of superego" and "integration of the ego." In a sense, each examiner is forced to define these phrases operationally in terms of test protocols and hope that they mean the same things in TAT material as they do in dreams, therapeutic sessions, etc. (*e*) While reference is made to experimental and clinical studies of TAT by other authors, these findings seem to have no observable impact on the question of the validity of Bellak's "system."

In all of the above, I am not presuming that results achieved with Bellak's methodology lack validity. I only mean to point out that the evidence for any judgment in this regard is almost entirely lacking. It seems to this reviewer that there should be a clear distinction made between theoretical constructs, admirable as they may be, and the *validity* of evaluative methods for describing or predicting specified aspects of behavior on the basis of thematic protocols, which are developed *in relation to* such constructs. It is conceivable that the first pillar may be essentially sound and that the second, inferred and operational, pillar may be of questionable soundness.

MAX L. HUTT
University of Michigan

COMMENT CONCERNING SAPPENFIELD'S REVIEW OF ARNOLD AND GASSON'S *THE HUMAN PERSON*

MAGDA B. ARNOLD
Loyola University

In the January number of the *Psychological Bulletin*[1] Dr. Bert R. Sappenfield wrote a review of *The Human Person*,[2] edited by Dr. J. A. Gasson and myself, in which he says:

> The reader is at once impressed by the fact that the authors . . . subordinate psychological principles to the articles of their religious faith—that they attempt to remodel psychology in the image of an ideology (p. 89).

And he closes:

> The non-Catholic reader . . . is likely to find it disappointing since it provides . . . little more than a discussion of psychological principles in terms of their agreement or disagreement with Catholic doctrines (p. 90).

Lest the reader take these comments to be a statement of fact, let me point out that they reflect a serious misunderstanding of our position as stated in the very first chapter of the book and adhered to throughout. In the first chapter we have tried to make explicit the various assumptions on which psychological theories can be based (e.g., that man is only quantitatively different from animals, or, that he has some capacities the animal does not possess). These assumptions either reflect the world view of the scientist (hence an unacknowledged philosophy) or they may be taken from a particular philosophical system. Our own assumptions are presented as an alternative to the currently popular world view (see Chap. 1, pp. 11–16)[2] and are based on Thomistic philosophy. This, like any other kind of philosophy, is the fruit of reflection, and hence has no intrinsic connection with Catholic doctrine. Non-Catholics or even non-Christians may be Thomists; one of the best known modern Thomists is, for instance, Mortimer Adler. In contrast, Catholic doctrine is based on revelation, is a matter of supernatural faith and makes no pronouncement on the matter or method of empirical or natural science. Hence Catholic doctrine does not touch the subject matter of psychology, namely the psychological activities of the human being.

Obviously, every philosophy can be called an "ideology:" the materialistic determinism on which Freud's system is based as well as the logical positivism that is the basis of much modern psychological theory[3] or the dialectical materialism underlying Russian psychology. In that sense Thomistic philosophy is an ideology, too, and Dr. Sappenfield is free to take it or leave it, as suits his own systematic position. But to imply that the authors of *The Human Person* attempt to "remodel psychology in the image of an ideology" while other psychologists do not, is to take unfair advantage of the philosophical naiveté of the reader. And to intimate that our particular systematic position cannot lead to valid scientific conclusions, and hence is little more than Catholic propaganda, betrays an unconscious prejudice that is surprising in a psychologist.

Another misrepresentation of our position is apparent in Dr. Sappenfield's statement that in our book

---

[1] *Psychol. Bull.*, 1955, 52, 88–90.

[2] MAGDA B. ARNOLD & J. A. GASSON, *The human person: an approach to an integral theory of personality.* New York: Ronald Press, 1954.

[3] Cf. M. H. MARX (Ed.), *Psychological theory: contemporary readings.* New York: Macmillan, 1951.

> the ego is conceived to function independently of, and to transcend the influence of, motivation. Thus, the authors deny determinism and subscribe to "free choice" or "free will"—a conception . . . that denies the potential predictability of human actions (footnote 1, p. 89).

But in the first chapter of our book we take great pains to show that self-determination is *not* independent of motivation but acts by eliminating alternatives and strengthening one particular motive. Moreover, throughout the book we point out that self-determination does not imply either uncaused action or unmotivated action or unpredictable action. On the contrary, self-determination is the factor that makes prediction possible (cf. p. 25, p. 40, p. 224).[2] Now it is the reviewer's right, if not his duty, to take issue with the logic of our argument if that is at fault; in that case, he has to point out the defect in the logic, not just deny the conclusion. But to mistake the authors' position and then attribute to them the bad logic consequent upon such misstatement is hardly responsible reviewing.

BOOKS AND MONOGRAPHS RECEIVED

ALLPORT, FLOYD H. *Theories of perception and the concept of structure.* New York: Wiley, 1955. Pp. xxii+709. $8.00.

BECKER, HOWARD, & HILL, REUBEN. (Eds.) *Family, marriage and parenthood.* (2nd Ed.) Boston: D. C. Heath, 1955. Pp. x+849. $6.25.

BRAM, JOSEPH. *Language and society.* Garden City: Doubleday, 1955. Pp. viii+66. $.95.

CALLIS, ROBERT, *et al. A casebook of counseling.* New York: Appleton-Century-Crofts, 1955. Pp. ix+352. $5.00.

CRONBACH, LEE J. (Ed.) *Text materials in modern education.* Urbana: Univer. of Illinois Press, 1955. Pp. 216. $2.50.

CROSSER, PAUL K. *The nihilism of John Dewey.* New York: Philosophical Library, 1955. Pp. xi+238. $3.75.

EDWARDS, ALLEN L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954. Pp. xvii+542. $6.50.

FERM, VERGILIUS. *A dictionary of pastoral psychology.* New York: Philosophical Library, 1955. Pp. xi+336. $6.00.

FRANK, LAWRENCE K. *Individual development.* Garden City: Doubleday, 1955. Pp. vii+52. $.85.

GILES, H. HARRY. *Human dynamics and human relations education.* New York: New York Univer. Press, 1954. Pp. vii+108. $1.25.

GILLESPIE, JAMES M., & ALLPORT, GORDON W. *Youth's outlook on the future.* Garden City: Doubleday, 1955. Pp. ix+61. $.85.

GREER, SCOTT A. *Social organization.* Garden City: Doubleday, 1955. Pp. x+68. $.95.

GRINGS, WILLIAM W. *Laboratory instrumentation in psychology.* Palo Alto: National Press, 1954. Pp. vi+282. $4.75 cloth. $4.00 paper.

HAVIGHURST, ROBERT J., & NEUGARTEN, BERNICE L. *American Indian and white children.* Chicago: Univer. of Chicago Press, 1955. Pp. xiii+335. $5.00.

HOIJER, HARRY. *Language in culture; conference on the interrelations of language and other aspects of culture.* Chicago: Univer. of Chicago Press, 1954. Pp. xi+286. $4.50.

HYMAN, HERBERT H., *et al. Interviewing in social research.* Chicago: Univer. of Chicago Press, 1954. Pp. xvi+415. $8.00.

KARN, HARRY W., & WEITZ, JOSEPH. *An introduction to psychology.* New York: Wiley, 1955. Pp. xi+315. $3.90.

KRAMER, JOSEFINE. *Intelligenztest.* Solothurn: St.-Antonius-Verlag, 1954. Pp. 288. s Fr. 29.

LAZARSFELD, PAUL F., & ROSENBERG, MORRIS. (Eds.) *The language of social research.* Glencoe: The Free Press, 1955. Pp. xiii+590. $6.75.

MACFARLANE, J. W., ALLEN, L., & HONZIK, M. P. *A developmental study of the behavior problems of normal children between twenty-one months and fourteen years.* Berkeley: Univer. of Calif. Press, 1954. Pp. vii+222. $2.25.

MCNEMAR, QUINN. *Psychological statistics.* (2nd Ed.) New York: Wiley, 1949, 1955. Pp. vii+408. $6.00.

MEEHL, PAUL E. *Clinical vs. statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. x+149. $3.00.

METZGER, WOLFGANG. *Psychologie.* Darmstadt: Dr. Kietrich Steinkopff, 1954. Pp. xix+407. DM 35.—

PIKE, KENNETH L. *Language; in relation to a unified theory of the struc-*

ture of human behavior. Glendale: Summer Institute of Linguistics, 1954. Pp. x+170.

PURCELL, THEODORE V. *The worker speaks his mind; on company and union.* Cambridge: Harvard Univer. Press, 1953. Pp. xix+344. $6.00.

RALLI, ELAINE P. (Ed.) *Adrenal cortex.* (Transactions of the Fifth Conference.) New York: Josiah Macy, Jr., Foundation, 1954. Pp. 187. $3.75.

REITMAN, FRANCIS. *Insanity, art and culture.* New York: Philosophical Library, 1954. Pp. 111. $3.75.

ROBINSON, M. F., & FREEMAN, W. *Psychosurgery and the self.* New York: Grune & Stratton, 1954. Pp. ix+118. $3.00.

ROTTER, JULIAN B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954. Pp. xvi+466. $6.50.

ROYCE, JAMES E. *Personality and mental health.* Milwaukee: Bruce Publishing Co., 1955. Pp. xiv+352. $3.50.

SEASHORE, STANLEY E. *Group cohesiveness in the industrial work group.* Ann Arbor: Univer. of Michigan Press, 1954. Pp. vi+107.

SMITH, WATSON, & ROBERTS, JOHN M. *Zuni law; a field of values.* (Reports of the Rimrock Project, Values Series No. 4.) Cambridge: Peabody Museum, Harvard Univer., 1954. Pp. ix+175. $3.00.

STONE, CALVIN P. (Ed.) *Annual review of psychology.* Volume 6. Stanford: Annual Reviews, 1955. Pp. vi+517. $7.00.

SUEHSDORF, ADIE. (Ed.) *Facts of life for children.* New York: Bobbs-Merrill, 1954. Pp. 92. $2.75.

ULETT, GEORGE. *Rorschach introductory manual.* Beverly Hills: Western Psychological Services, 1955. Pp. 45. $3.50.

THE WOODS SCHOOLS. *The Adolescent exceptional child.* (Proceedings of the 1954 Spring Conference of the Child Research Clinic of the Woods Schools.) Langhorne: The Woods Schools, 1954. Pp. 78. No charge.

ZANGWILL, O. L. *Psychology as the study of behavior.* New York: Cambridge Univer. Press, 1955. Pp. 36. $.50.

ERRATUM

In the article entitled "Estimating the chance expectancies of diadic relationships within a group" by R. Tagiuri, J. S. Bruner, and N. Kogan (this JOURNAL, 1955, **52**, 122–131) there is an error in the second column of page 124. The sentence should read as follows:

"Certain diads can occur in only one way, e.g., $S_i\!\rightarrow\!\leftarrow\!S_j$; others can come about in two ways, e.g., $S_i\!\rightarrow \qquad S_j$ could also be $S_i \qquad \leftarrow\!S_j$."

There is also an error in the last sentence of footnote 3 on page 126. The sentence should read as follows:

"Thus $S_i\!\rightrightarrows\!\dashleftarrow\!\!\dashleftarrow S_j$ is *not* symmetrical, while $S_i\!\dashrightarrow\!\!\dashrightarrow\!\leftleftarrows S_j$ is."

VOL. 52, NO. 5 SEPTEMBER, 1955

Psychological Bulletin

CONTENT-ANALYSIS STUDIES OF PSYCHOTHERAPY

FRANK AULD, JR. AND EDWARD J. MURRAY[1]
Yale University

It is only sixty years since Freud developed psychoanalysis (**40**, p. 252) and thereby originated dynamic psychotherapy. Although psychoanalysts and other investigators have learned a great deal about psychotherapy in these sixty years, studies of psychotherapy have suffered from three hindrances:

1. The basic data of psychotherapy were transient. Furthermore, they were accessible only to the therapist; others had to take his word for what happened in the interviews. The consequences of this lack of adequate recording are eloquently set forth by Kubie (**44**).

2. Conclusions stemming from investigations were matters of impression and opinion, because there was no technique for studying the verbal materials objectively.

3. The data could not be fitted into a suitable theoretical framework because the psychologies of the day (for example, Wundt's and Brentano's) had little relevance to the phenomena observed. To appreciate this, one might imagine oneself in Freud's place in 1895 and ask, "How could I explain the things I've observed?"

Recent methodological and theoretical developments seem to justify the hope that these three obstacles to scientific research on psychotherapy can be overcome. The new methods are sound recording and content analysis; the new theoretical developments are the recent attempts to develop a general science of behavior.

Sound recording of interviews has made a common set of data available to scientists, a set that can be preserved and be studied as many times as necessary (**7, 26, 76**). Pioneers in the recording of psychotherapy interviews include Zinn, Lasswell (**45, 46, 47**), Rogers (**80, 81**), Robinson (**79**), Covner (**14, 15, 16**), and Porter (**70, 71**). All of the studies reviewed in this article, with one exception, derived their data from sound recordings.

Content analysis is a method for studying the content of communication in an objective, systematic, and quantitative way (Berelson, **4**, p. 18). "Content" is what is said. Lasswell (**48, 49**) pioneered the application of content analysis to social science problems. In the past decade content analysis has been widely used for studying therapeutic interviews. All the studies reviewed in this paper are content-analysis studies.

---

[1] The authors wish to thank Professor John Dollard for his guidance and support of their research. Professor Dollard read this paper and made helpful suggestions. This paper is a product of research project M-648, "Development of quantitative methods for detailed study of psychotherapy," supported by the National Institute of Mental Health, U. S. Public Health Service.

General theories of behavior, which have been developed in recent years, provide possible frameworks for explaining the data of psychotherapy and for guiding investigators toward studies that will advance our understanding of human behavior as well as give answers to specific questions (**32**). The fusion of psychological theory with psychotherapeutic research is exemplified in the work of Dollard and Miller (**20**). Although the nondirective group at first had no comprehensive theoretical platform, in recent years these investigators have shown a growing tendency to state explicit theories and to design their investigations to test hypotheses derived from the theories (**83**, **84**). And, of course, it cannot be forgotten that the psychoanalysts, building on their clinical experience, constructed a comprehensive psychological theory. Despite its vagueness and lack of rigorous formulation, it has served as a guide for a number of investigators.

In this paper we attempt to review the considerable body of literature on content analysis of recorded interviews.

A SURVEY OF THE STUDIES

The content-analysis studies of psychotherapy fall into three general classes: (*a*) Methodological studies, in which the aim was principally to develop measures, (*b*) Descriptive studies of cases, and (*c*) Theoretically guided studies of therapy (i.e., studies of cause-and-effect relationships). Although it is admitted that the third kind of study, if well done, contributes the most to our understanding, studies of the first type are necessary to develop methods and studies of the second type may provide hunches that can be rigorously tested later by a theoretically oriented study.

Methodological Studies

In these studies, the emphasis is on development of new measures rather than on testing of hypotheses. Nevertheless, the inventor of a measure always has some theoretical presuppositions; and these are what influence him to measure one kind of thing rather than another. Porter (**70**, **71**), for instance, developed a classification of therapist responses. His classification emphasizes the degree to which the therapist takes responsibility for the interview, which Porter considers an important variable in therapy (**72**, pp. 45–60). Among the nondirective therapists, Royer (**87**), Snyder (**91**), Curran (**17**), Raimy (**73**), Hogan (**34**), Haigh (**31**), Stock (**93**), Hoffman (**33**), Kahn (**41**), and others have developed measures of various kinds. We shall choose several measures for special comment, because they illustrate the problems involved in such classifications.

Snyder's system. Snyder's system (**91**) has been especially influential. Seeman (**88**), Aronson (**1**), Rakusin (**74**), Tucker (96), Gillespie (**27**), and Blau (**6**) have made use of it in their investigations. In Snyder's system, the therapist's responses are classified according to a modification of Porter's categories. The categories designate the technique which the therapist uses: restating content, clarifying feeling, interpreting, structuring, leading, suggesting, questioning, persuading, accepting, reassuring, approving, disapproving. The client's responses are classified under these headings: problems, simple responses (questions, answers, acceptance, disagreement), insight, planning. Notice that Snyder, like Porter, considers the degree of responsibility assumed by the therapist to be an important variable in the therapist's behavior. On the side of the client, Snyder's classification implies that it is important to notice whether the client is discussing problems or showing understanding. Presumably in successful cases problems dominate at the beginning and insight and plans dominate at the end.

Curran's two systems. Curran (**17**) has reported a detailed analysis of a single case. His study is of special interest for two reasons. First, he measured "insight" by noting instances of the client's connecting two different problems. Also of interest is Curran's classification of the problems discussed by the client. The problems discussed by the client included: hostility, dependency, insecurity, unhappiness, conflict, discouragement, withdrawal, daydreaming, feelings of inferiority, sex, sin, younger brother, school work, and war. The special interest of this classification lies in the fact that while most researchers have only noted whether the client talked about a problem or about something else, Curran described *what* the problems were that the client talked about.

Haigh's measure of defensiveness. Haigh (**31**) developed a measure of defensiveness, building on previous work by Hogan (**34**). Haigh defined defensive behavior as behavior that shifts attention away from an inconsistency that, if perceived, would threaten the client.[2] It may be noted, by the way, that in order to judge a remark as defensive according to Haigh's system one must make the judgment that two of the statements of the client are incongruent or that the client "ought" to be exploring some aspect of his emotional life and is avoiding it (**31**, p. 182).

Interaction Process Analysis. The theory behind this system, which was developed by Bales (**3**), is that problem solving involves a series of steps: getting information, making decisions, carrying out actions. At each stage of problem solving, the participants interact with each other. One participant, for instance, asks for information; another gives information; then a third member of the group may offer an opinion of the correctness of this information; the individual who gave the information may then proceed to defend his views with some heat. As applied to therapy, the Bales system is used to score both the remarks of the therapist and those of the client.

Discomfort-Relief Quotient. The Discomfort-Relief Quotient, or D.R.Q., was suggested by behavior theory. Behavior theory states that responses are incited by drives and reinforced by drive reduction. The learning of new habits is, according to this theory, accompanied by drive reduction. The new learning that occurs in successful therapy ought, therefore, to be accompanied by a reduction in drive. Guided by these theoretical notions, Dollard and Mowrer (**21**) attempted to measure the amount of drive borne by the client. They did this by classifying each word as a drive, reward, or neutral word according as the word seemed to represent discomfort (suffering, tension, pain, unhappiness) or relief (comfort, satisfaction, enjoyment) or neither. The D.R.Q. is obtained by dividing the number of discomfort words by the total number of discomfort *and* relief words. The authors demonstrated very satisfactory reliability of D.R.Q. scores. Dollard and Mowrer, employing other units besides words, found that sentences and "thought units" could also be reliably scored.

---

[2] This definition seems essentially the same as the psychoanalytic definition of "resistance." Colby, for example, defines resistances as "those defenses that operate in and against the therapeutic process to prevent an uncovering and a dissolution of the neurotic conflict" (**12**, p. 8). Haigh's definition differs from the analytic one, however, in this respect: He notes only those resistances that can be easily inferred from the content of the client's speech, i.e., from the client's denials, rationalizations, and projections. Analysts, on the other hand, note not only these but also such evidences of resistance as: silences, avoidance of obvious topics, lateness in keeping appointment, and various transference responses (**12**, pp. 95–106).

Dollard and Mowrer intended the D.R.Q. as a measure of the tension *experienced* by the client. We know, however, that the client's verbal responses do not always accurately label his drives (**20**). It is possible that: (*a*) A verbal response may indicate a drive when the drive is not present, as when a client makes an insincere complaint in order to enlist the sympathy of the therapist. (*b*) A drive may be present with no verbal response describing it, as when a client denies having sexual wishes while his other behavior convinces us that he does have them. (*c*) A verbal response may describe a conflict different from the one the patient is actually experiencing, as when talk about vocational problems replaces talk about an unconscious homosexual conflict. It is also possible that the importance of a drive is not accurately reflected by the *frequency* of occurrence of sentences about it. It is the task of empirical research to discover whether clients accurately label their drives or whether these possible distorting factors interfere with accurate labeling to such a degree that verbal responses cannot be used as indices of drive.

In a study of interviews from six recorded psychotherapy cases, Mowrer *et al.* (**62**) found a moderately high relationship between the D.R.Q. scores for interviews and measures of palmar sweating made after the same interviews. For five of these cases, the *r* between D.R.Q. and palmar sweating was .57; in the sixth case, for which the correlation was computed separately, the *r* was only .30. To the extent that palmar sweating is itself an adequate measure of "tension," this finding somewhat strengthens the case for considering the D.R.Q. a measure of tension. Meadow *et al.* (**53**), studying psychotic patients, found no significant relationship between D.R.Q. score in a special interview and a psychiatrist's rating of tension made on the basis of other interviews with the patient. In interpreting this result one should note, first, that the rating of tension was not made in the same situation in which the D.R.Q. score was obtained; second, that the result of a study of psychotic patients, who as a group are likely to make verbal responses without showing "appropriate" emotional reactions, may not predict very well what will be found when normal and neurotic subjects are studied; and, finally, that the validity of the psychiatrist's ratings is not known.

The problem of how the verbal behavior of the client, which is taken note of in content analysis, is related to his nonverbal behavior, is a difficulty not only for the D.R.Q. but also for every other system of content analysis.

Positive-Negative-Ambivalent Quotient. A measure independently developed by Raimy (**73**) has a strong family resemblance to the D.R.Q. Raimy's PNAvQ is a quotient, as the D.R.Q. is; and, like the D.R.Q., it is an indication of positive and negative emotional reactions of the client. It differs from the D.R.Q. in focusing attention on the statements of the client about himself (whereas the D.R.Q. includes all statements of the client, whether they are self-referential or not). In some cases, the two measures would be expected to give similar results, which is what Kauffman and Raimy (43) found to be true in a study of 17 counseling interviews. These authors also found the PNAvQ to be somewhat lower, on the average, than the D.R.Q.

Horwitz (35), however, found only a low correlation ($r = .38$) between the D.R.Q. and PNAvQ in 36 initial casework interviews. It is possible that the low correlation should be attributed to unreliability of scoring; but this seems unlikely, since investigators using these measures in other studies have found them to be reliable. It is more likely that the correlation is low because the D.R.Q. includes all statements of the client, whether about himself or about others, while the PNAvQ includes only statements referring to the client. If the client talks a great deal about other persons, these statements will affect the D.R.Q. but not the PNAvQ—with a resulting drop in agreement between the two measures.

Doubting that a measure of the "self concept" would be much related to the D.R.Q. if the "self concept" were narrowly defined, Horwitz also investigated the correlation between the D.R.Q. and a "self-approval quotient." The self-approval quotient is the same as the PNAvQ except that statements which indicate that the client is happy, glad, or improving, or fearful, anxious, or unsuccessful are not included. (Such statements *are* included in the PNAvQ.) The only statements included in the self-approval quotient are those in which the client evaluates himself. The D.R.Q. was not significantly correlated with the self-approval quotient ($r = .14$).

A system in terms of motivation and conflict. E. J. Murray (63, 64), in devising a system for the study of motivation and conflict in psychotherapy, was influenced by psychoanalysis and learning theory. Murray wished to designate underlying motivations of the client; but, at the same time, he wanted the system to be objective, i.e., to have explicit rules for inferring the underlying motives. In keeping with the wish to have an objective system, Murray chose categories that require relatively little inference by the scorer and reflect chiefly manifest content. The category system designates statements expressing a need, statements expressing anxiety about a need, and statements expressing hostility on account of frustration of a need. The drives included are: sex, affection, dependence, independence, and "unspecified" drive. The system also requires the scorer to designate the person who is the object of the need, e.g., the person who is loved or hated.

Murray (63, 64) also devised categories for describing the therapist's behavior, building on the theoretical work of Dollard and Miller (20). Remarks of the therapist were classified according to activity-passivity and according to their function as rewards, punishments, labels, discriminations, generalizations, instructions concerning free association, directions, or probes.

The reliability of scoring these categories was studied (63, 64) and found to be fairly high. It was found that reliability decreased as a greater number of simultaneous discriminations were required of the scorer. This finding agrees with the experience of Kaplan and Goldsen (42).

Lasswell's general purpose system. Lasswell (48) has suggested a scheme for classifying interview materials under very general headings, so that the system can be applied to interviews differing widely in topic. The classification tells who makes the statement (talker-listener-another), whom the statement refers to (talker-listener-another), and what attitude is expressed (favorable-unfavorable). To our knowledge this scheme has not been applied to psychotherapy cases, except apparently in a recent study by Rosenman (**85**).

Other measures. Other proposed measures include: Collier's (**13**) scale for the degree of "uncovering" used by the therapist, Helen E. Miller's (**54**) measure of "acceptance," Raskin's (**75**) "locus-of-evaluation" rating, Elton's (**22**) responsibility rating, Carnes and Robinson's (**9**) talk ratio, the various measures used by Lewis in her pioneering study (**50**), and White's (**97**) frustration-satisfaction ratio, a measure similar to the D.R.Q.

Descriptive Studies

Movement from problems to insight. Snyder (**91**) has described how the content of the client's speech changes during the course of nondirective therapy. According to Snyder, first there is a statement of a problem or problems; then discussion of these problems, with increased insight; and finally, formulation of plans for new responses. Studying four "successful" cases and one "unsuccessful" case, Snyder (**92**) found that the successful cases showed a trend from problems to plans; the unsuccessful case failed to show this trend. Seeman (**88**) repeated Snyder's study, using a larger sample of cases, and confirmed his results. Results reported by Snyder (**91**), by Seeman (**88**), and by Blau (**6**) also indicate that in the cases that were believed successful, the client talked less about problems and more about insights and plans at the end of treatment than at the start; similarly, he expressed more positive feelings and fewer negative feelings than he did at the beginning. Such a change was not observed in unsuccessful cases. This result is hardly an independent confirmation that the cases judged to be successful *were* successful, since the judges undoubtedly reacted to just such changes in the client's verbalizations when deciding whether cases were successful. While these studies do not give independent confirmation, they are nevertheless valuable because they specify some ways in which the client's speech behavior changed, and they provide reliable measures for these changes.

Attitudes toward self. Raimy (**73**) studied the kinds of attitudes toward self expressed by the client at various times in the course of therapy. He found that in cases that were judged to be successful, positive statements about self increase and negative statements decrease during the course of treatment. Seeman (**88**) studied all attitudes of the client, both about self and about other people, and found that positive expressions increase and negative expressions decrease during therapy. He found no significant correlation between the change from negative to positive attitudes and the counselor's rating of success of the case, if all statements of attitude were considered; but when he examined only statements expressed in the present tense, he found a significant correlation between changes in these statements and the counselor's rating ($r = .66$). Bugental (**8**) has used Raimy's scales and somewhat modified them.

Studies using the D.R.Q. Still another way of describing the course of psychotherapy is to note changes in the degree of tension expressed in the client's sentences. A measure of verbal tension (the D.R.Q.) has already been described in the preceding section, and the theoretical considerations that guided its invention were presented there. Results with the use of the D.R.Q. will be briefly presented here. Hunt (**36, 37, 38, 61**) found little agreement between changes in the D.R.Q. (scored from caseworkers' summaries) from the beginning to the end of casework and the degree of movement or progress in the case as judged by experienced caseworkers. When these judgments of experienced caseworkers were quantified and made highly reliable, by use of a scale of "movement," there was still little agreement between movement and drop in D.R.Q. Assum and Levy (**2**), studying one case, found a drop in D.R.Q. accompanying success. Natalie Rogers (61) studied three cases, one judged as very successful, one as moderately successful, and one as unsuccessful. The successful cases showed a drop in D.R.Q., but the unsuccessful one did not. Cofer and Chance (**10**) computed the D.R.Q. for each hour of five published cases—four nondirective cases and the case in Lindner's *Rebel Without a Cause* (**51**). All the cases were judged as successful by the therapists, and all showed a drop in D.R.Q. Mowrer (**61**) reported one of his own cases which showed a drop in D.R.Q. and was believed to be a successful case.

On the evidence so far, there is no clear relationship between the change in D.R.Q. from beginning to end of a case and the success of the case. Some investigators have reported a drop in D.R.Q. accompanying success (these investigators studied, in total, verbatim transcripts of 10 cases); others (who studied case workers' notes of 38 cases) found no relationship between drop in D.R.Q. and success of the cases. It may be noted that no evidence for the reliability of the judgments of "success" was reported for those 10 cases showing positive results; but reliability of the judgments of "movement" in the 38 cases was established.

Murray, Auld, and White (**65**) have reported a "partly successful" case showing no drop in D.R.Q. Their paper includes a discussion of the reasons for absence of any change in the D.R.Q.

The critical issue here, it seems to the present authors, is that we have no adequate measure of "success." Until such a measure has been devised, it makes little sense to ask whether a drop in D.R.Q. is related to success.

Relations of problems to each other. Curran (**17**), studying 20 recorded interviews of a single case, described changes in the client's perceptions of relationships between his problems. At the beginning of therapy, the client discussed a large number of problems and talked about them separately; at the end of therapy he discussed a smaller number of problems, tending to combine problems that had previously been separate. Curran's contribution to methodology is his idea of studying "insight" by noting the client's discovery of relationships between previously separate problems.

Differences between different therapies. Porter (**70**, **71**) described differences between "directive" and "nondirective" counselors. The directive counselors studied by Porter talked much more than the nondirective counselors and took more responsibility for guiding the interview by questioning, explaining, probing, and advising. Gump (**30**) compared responses of the therapists in one psychoanalytic and five nondirective cases and noted differences in technique. No generalizations about analytic or nondirective therapy should be made, however, on the basis of only six cases.

Strupp (**94**) has compared the verbal responses of Rogerian and non-Rogerian therapists when they were asked what they would say after various statements by a client. Answers of the therapists were classified according to Bales' system, with fairly high reliability (78% agreement between two scorers). As would be expected, there were differences between the answers of Rogerian and non-Rogerian therapists. Variation within the two groups of therapists was also large. Strupp (**95**) further classified the non-Rogerian therapists according to whether they had received a personal analysis. Those who had been analyzed were more likely to give answers classified as "disagrees" and "shows antagonism" and less likely to give answers coded as "agrees, shows passive acceptance." The analyzed therapists less often indicated that they would say nothing at all after the client's statement.

Group psychotherapy. Coffey *et al.* (**11**) have presented a description of the content of group psychotherapy interviews with college students. The nuclear conflict of students in these groups was described by the authors as follows: "High standards and intellectualized ideals are often associated with passivity and inhibition of emotions. The constricted feelings find distorted expression in intellectualization and isolated fantasy resulting in varying amounts of guilt, immobilization, and affective sterility" (**11**, p. 58). A classification of what the groups talked about yielded the following results: Sexuality, 27% of total time; vocational problems, 14%; attitudes toward psychotherapy, 12%; attitudes toward and perception of self, 11%; society and the individual, 10%; interpersonal and social relations, 9%; relations to authority, 6%; family relations, 6%; handling of hostile feelings, 4%.

Roberts and Strodtbeck (**78**) applied the Bales system (Interaction Process Analysis) to group therapy interviews of depressive and schizophrenic patients. They expected that the paranoid schizophrenic patients would use relatively more of the negative responses ("shows aggression" and "disagrees") and would make fewer responses to each other and more to the leader. The depressed patients were expected to make fewer responses per minute than the schizophrenic patients. All three of these predictions of differences between the groups were borne out by the data.

Gorlow *et al.* (**28**) studied nondirective group therapy sessions, using a system of categories similar to that developed by Snyder.

Application of Bales system to student counseling. Perry and Estes (**69**), with the collaboration of Bales, applied the Interaction Process Analysis system to four counseling interviews with a student. In this counseling case, the therapist shifted from responses categorized as "gives orientation" to responses falling in the categories "gives opinion" and "asks for opinion." The student showed a falling off, during the four interviews, in the percentage of responses classified as "shows tension, asks for help, withdraws out of the field." The authors interpreted the shift in therapist responses as documenting their belief that the counselor used nondirective techniques at the start of counseling, then shifted to responses that defined the counseling process as a collaborative venture. The drop in the client's "asks for help" responses was interpreted as showing his acceptance of a collaborative view of therapy and his giving up expectations that the therapist would play an authoritarian role.

Studies of linguistic characteristics. It is beyond the scope of this paper to discuss here studies of the grammatical structure of clients' speech during therapy. Pertinent studies include those of Roshal (**86**), Zimmerman and Langdon (99), and Grummon (29). These studies and others have been reviewed by Mowrer (60).

Theoretically Guided Studies

Lasswell's pioneering work. Lasswell was a pioneer not only in making sound recordings of analytic interviews, but also in applying content-analysis methods to their study. His earliest report on the analysis of transcribed interviews (45) indicates that he classified the client's utterances according to whether they referred to the interviewer or not. He proposed the hypothesis that *conscious affect* is indicated by references to the interviewer and that *unconscious tension* is indicated by slow speech, pauses, etc. If this hypothesis is correct, then tension as shown by slow speech should be correlated with other measures (e.g., physiological measures) of tension. In the cases studied Lasswell found a correlation between verbal indices of tension (slow speech) and physiological indices (high conductivity of the skin). On the other hand, conscious affect (indicated by references to interviewer) was correlated with increased heart rate. It is not clear to us why this latter relationship should have been found. Lasswell, however, interpreted it as supporting his hypothesis concerning the meaning of references to the therapist. The measures of unconscious tension (slow speech and high skin conductance) were negatively correlated with the measures of conscious affect (references to therapist and increased heart rate).

In another study (47) Lasswell obtained similar results. In this later study he included measures of blood pressure before and after each hour and of gross bodily movements during the hour. He found that unconscious affect (defined as above) decreased in the course of the interviews and conscious affect increased, as would be expected to occur in successful psychoanalytic treatment.

Is clarification of feeling the best technique? Rogers (82) has asserted his belief that progress in therapy always follows "recognition of feeling" by the therapist. Analytic therapists take a quite different view, believing that interpretation of resistance and labeling of previously unlabeled emotions are essential therapeutic techniques (Freud [23, 24], Dollard and Miller [20], Colby [12]). The available evidence from content-analysis studies concerning effectiveness of various techniques of the therapist is summarized in this section.

Snyder (91) studied the client responses following various kinds of responses by the therapist. He found that "insight" and "discussion of plans" by the client were more likely to follow nondirective than directive responses of the therapist.

Sherman (90) in a study of student counseling interviews found that "tentative analysis" by the therapist was most likely to be followed by client responses rated high on her "working relationship" scale; "interpretation" and "clarification" were moderately likely to be followed by good "working relationship;" and "urging" was the therapist technique least likely to be followed by good "working relationship." In discussing her study Robinson (79, p. 130) points out, however, that "a primary technique can have just about every degree of effect. Part of this range is due to the unreliability of rating scales, but in great part this range shows that at times each technique was a highly effective way of handling a particular unit or, in the case of poor outcomes, that the particular technique was the wrong one to use or that, with the interview conditions as they were, no particular technique could have worked." Sherman also investigated the effect of technique used by the therapist on degree of insight shown by the client in the following response. In "adjustment-problem" units, clarification was most likely to be followed by a response rating high in "insight;" "interpretation" and "tentative analysis" were next most likely to be followed by "insight;" and "urging" was least likely to be so followed. It should be noted that the "insight" spoken of here is *insight-as-judged-by-Snyder's-or-Sherman's-judges*—and that other investigators might not have agreed with these authors concerning what is or is not insight. Psychoanalysts, for example, might require the patient to discover unconscious motivations before crediting him with the achievement of insight. Snyder and Sherman were dealing with something other than this.

In another study bearing on this point, Bergman (**5**) reached the conclusion that "reflection of feeling" was the only technique that led to insight or continued exploration. According to Bergman, interpretation by the therapist led to the abandonment of self-exploration. Contradictory results were obtained by Gillespie (**27**), who found that "verbal signs of resistance tend to be proportional to the number of counselor statements regardless of the counselor category" (**27**, p. **119**). Hostile expressions toward the therapist and therapy were more likely to occur after interpretation or "inaccurate clarification of feeling" by therapist than after "restatement of content" and "accurate clarification of feeling." Other signs of "resistance" noted by Gillespie, however, were not significantly associated with interpretative activity by the therapist. These other signs included: client-initiated long pauses, short answers, stereotyped repetition of the problem, changing the subject being discussed, excessive verbalization and intellectualizing, and emotional blocking. These signs of "resistance" are exactly those which psychoanalysts take as indicating resistance (**12**, pp. 96–98). Analytic therapists, however, apparently do not worry so much about the client's attacks and disagreements as the nondirective therapists do, except when these client responses threaten to interrupt communication.

In a study of a single case, Dittmann (**18**) noted what kinds of therapist responses were likely to be followed by client behavior that was judged to indicate progress. Client responses indicating "progress" were most likely to come after therapist responses that were slightly more interpretive than pure "reflection."

What can we learn from these studies? We learn, especially from the work of Robinson and his students, that there is no single type of response by the therapist that works best under all circumstances. Effectiveness of a response by the therapist probably depends on the expectations of the client (e.g., whether he expects the therapist to give advice or not), on the particular circumstances of the situation, and on the client's ability to tolerate increased self-knowledge at that particular time. We also learn that judgment about the effectiveness of a technique depends on having some measure of effect and that different measures of "effect" yield different judgments about what produces effectiveness. Measures of effectiveness (e.g., of "insight") that are considered appropriate by nondirective therapists may not necessarily be considered appropriate by analytic therapists—and vice versa.

Differences in therapist responses. Reid and Snyder (77) investigated the differences between counselors in their responses to identical client statements. Phonographically recorded client statements were replayed to a group of 15 clinical psychologists. After each statement the psychologists were allowed 15 seconds to write down the principal feelings they believed the client had expressed. Reid and Snyder found, first, that the same client response evoked a variety of responses from the counselors. They reported, further, that the counselors who were judged by their instructors to be more skillful were more likely to label the client's feeling in the way that a majority of the group labeled it. The authors also observed that some counselors consistently tended to label clients' feelings in a preferred way, e.g., some counselors were likely to perceive the clients as reporting feelings of insecurity, while other counselors labeled the same client responses as indicating hopefulness and ambition. This study shows a promising way to apply experimental methods to the study of the therapist's behavior.

Haigh's study of defensiveness. Haigh (**31**), believing that "defensiveness" (incorrect perception of one's behavior) should decrease during successful psychotherapy, identified instances of defensive behavior in 10 client-centered therapy cases, counting the number of these defensive responses in each interview. Haigh found that the clients had fewer defensive responses in the second half of treatment than in the first half. Seven of the clients had fewer defensive responses during the second half; two had more defensive responses during the second half; one had no identified defensive responses during either half. The present authors have applied the exact test of significance to these data, and find that the first and second halves of therapy do not differ significantly ($p=.30$).[3] A larger number of subjects must be studied before a definite conclusion concerning the course of defensiveness throughout therapy can be drawn.

A further aim of Haigh's study was to show that awareness of defensiveness caused a decrease in defensiveness. According to Haigh, the client would at the beginning of therapy present a number of defensive opinions. As the nondirective therapist reacted to the client with acceptance, lack of moral evaluation, and willingness to have the client make his own decisions, the client would have progressively weaker motives to present defensive opinions. He would become aware of inconsistencies in his opinions, actions, and emotions; in the benign atmosphere of therapy, he would be able to abandon these inconsistencies.[4] According to this account of therapy, increased awareness of defensiveness should be followed, in successful therapy, by a decrease in defensiveness. Haigh did not make the detailed analysis that would be necessary to test the adequacy of this construction, but he did report whether awareness-of-defensiveness showed any change, on the whole, throughout therapy. Haigh found that defensive behavior of which the client was unaware declined in the six cases that showed a drop in total defensive behavior; "unaware" defensive behavior increased in the three cases that showed an over-all increase in defensive behavior. Remembering that these results are derived from a very small sample of cases, we can make the tentative interpretation that clients in nondirective therapy who are unaware of their defensive behavior cannot change it; clients who become aware of it, can and do change it. We would like, however, to see the process of increased awareness studied in greater detail.

---

[3] Haigh's statistical analysis, which resulted in a significant value of chi square, is inappropriate. The chi-square test requires the assumption that the items included in the table are independently selected. But the responses have not been independently selected from a population of responses; instead, if one response of a particular client was included in the sample, all his responses were included.

[4] Note the parallelism between this account of therapy and the account given by Dollard, Auld, and White (**19**). Dollard *et al.* believe, however, that the therapist must react with doubt and incredulity to elements of the client's account that do not make sense.

A study of hostility and defenses. An exploratory study making use of his content-analysis system has been reported by Murray (63). The hypotheses tested in this study were suggested by N. E. Miller's work on learning theory (55, 57, 58, 59, 67) and by psychoanalytic theory. Murray found that in the course of the psychotherapy case studied, defensive sentences of the client decreased and hostile sentences increased. He also found that interviews which contained a large number of defensive sentences had a relatively small number of hostile sentences ($r = -.73$). Intellectualization and physical complaints comprised the defensive sentences; statements of frustration and resentment comprised the hostile sentences.

The results were interpreted in terms of learning theory as follows: The defenses were assumed to be motivated by anxiety. When the client uttered a defensive sentence instead of a hostile sentence, he thereby escaped the anxiety which would have been aroused by utterance of the hostile sentence. During the psychotherapy the therapist interpreted one form of defense (physical complaints), while maintaining a permissive attitude toward the expression of hostility. These acts of the therapist resulted in a weakening of the defenses and extinction of the anxiety motivating them and thereby permitted an increased expression of hostility.

Murray also tested the hypothesis that the various hostile sentences would appear in a sequence which might be predicted by the theory of displacement (56, 58). His expectation was that the client would first express hostility about the less significant persons in his life and would gradually proceed to express hostility toward more important persons—for example, his mother. An examination of the data showed a sequence in the expression of hostility toward various persons, but not the sequence that had been predicted. Hostility was expressed first toward the more significant persons, then toward less significant persons. This finding led to a re-examination and reformulation of the displacement theory. Murray and Berkun made certain extensions of Miller's displacement theory and verified them in an animal experiment (66).

It was also discovered that the two defenses, intellectual discussion and physical complaints, seemed to operate as alternative members of a habit-family hierarchy. When one defense declined, the other tended to rise. Additional evidence for the hypothesis that the defenses functioned as members of a habit-family hierarchy is given by the fact that in the later stages of therapy, when anxiety was increasing following the greatly increased expression of hostility, the previously uninterpreted defense (intellectualization) increased much more than the defense that had been interpreted.

Attitudes toward self and others. Researchers of the nondirective group have investigated the changes in attitudes toward self and toward others that occur during psychotherapy. They believe that positive attitudes toward self and others are desirable, and they have been interested to discover what actions of the therapist can cause a change in these attitudes. One hypothesis is as follows: The therapist's warm, accepting attitude toward the client enables the client to recognize his own wishes more fully and evaluate his own behavior with greater acceptance (Rogers, **83**, p. 41). A second hypothesis is this: To the degree that client has positive feelings about himself, he can have positive feelings toward other people. A third hypothesis is this: A person who feels affirmatively about other people gets along with them better than a person who reacts negatively. From these three hypotheses it follows that the therapist by a warm, accepting attitude toward the client can cause the client to feel positive about other people and to get along better with them (Rogers, **83**, pp. 160, 520).

So far as we know, there has been no content study bearing on the first hypothesis. The second hypothesis was tested by Sheerer (**89**) and by Stock (**93**), who both found a correlation between positive opinions about self and positive opinions about others. The third hypothesis was investigated by McIntyre (**52**), who found no evidence that persons with high scores on an "acceptance-of-others" questionnaire were themselves accepted by others (as judged by a sociometric questionnaire). One must conclude that these three hypotheses have not yet been proved.

Prognostic studies. Can outcome of therapy be predicted from early clues in the case? There seem to be only two content-analysis studies bearing on this question. Lasswell (**46**) investigated whether certain measures of speech and of physiological responses, made in early hours, could be used to predict the success of treatment. He reported that a combination of measures taken during the first 10 or 12 hours permitted accurate prediction of the client's progress. Lasswell reported of his subjects: "Those who adjusted themselves readily to the situation (and who made progress in insight) showed rising skin resistance curves. Our interpretation is that they were finding satisfactions in the interview experience against which they did not maintain a strong inner defense. Unconscious tension was diminishing, even though active conscious affect might rise" (**46**, pp. 246–247). Since the number of cases studied by Lasswell was small, this promising exploratory work needs to be repeated with larger samples.

The other study of prognostic indicators is that of Page (**68**), who found that measures of variability of the content and feeling of the client's speech in the first interview were not correlated with outcome of treatment. The amount of talk by the client in the first interview had a small ($r = .30$) but statistically significant correlation with the criterion of success. It should be noted that Lasswell, Page, or any investigator who tries to find prognostic signs is hampered by the lack of any fully adequate definition or any adequate measure of "success" of therapy.

CHOICE OF A CONTENT-ANALYSIS SYSTEM

In choosing a content-analysis system, one faces all the problems of choosing a topic that deserves scientific investigation *plus* the problems involved in choosing appropriate methods after one has decided on the topic. A few comments will be made on each of these tasks.

Experience has shown that the most fruitful scientific investigations are those that bear some relation to a general theory or, at least, to a well-thought-out hypothesis (Wilson, **98**, p. 2). What hypotheses, then, might best be investigated? A large amount of attention has been focused on the immediate effects of various techniques used by the therapist and on the positive and negative opinions of patients. Little or no study by content-analysis methods has been made of such questions as these: What is the function of transference responses in psychotherapy? What is the effect of the therapist's giving love to the client? What is the role of childhood memories—must they be recaptured in successful therapy? How does the client learn new verbal units? What cues in the client's behavior does a good therapist respond to? These questions are relevant to a general theory of psychotherapy and thus, in our opinion, are worthy of study.

Once the investigator has chosen his field of inquiry, he must select appropriate techniques. If there are extant content-analysis systems which bear on his topic of inquiry, how can he decide whether to use one or more of them? This is the problem of validity. If the investigator is interested in the emotions of the client, he will want to know whether the D.R.Q. measures emotion. If he is interested in the influence of the therapist on the client, he may ask, "Does Porter's classification of therapist responses adequately measure this influence?" If he wants to assess therapeutic progress, he might ask, "Do measures of 'self-attitude' such as Raimy's PNAvQ really reflect better adjustment of the client? Do these measures neglect unlabeled emotional conflicts? Do they give equal weight to changes in fundamental psychological conflicts and to resistant escape sentences?"

Janis (**39**) has proposed that the validity of a content system be judged by the number of relationships found by investigators who have used the system. As Janis points out, this does not mean that a valid measure is related to every other valid measure. For instance, intelligence may not be very important in determining the success of psychotherapy. If such be the case, we will not call the Wechsler-Bellevue Scale invalid if it fails to correlate with measures of success in therapy. But the intelligence-test score *should* be related to other measures of intelligence, e.g., to success in school work. Similarly, content measures are valid if they are correlated with other measures, when we have reason to believe they *should* be correlated. Admittedly, it is hard to designate which variables ought to be related to each other, and so it is difficult to determine when a lack of correlation implies lack of validity.

In our opinion, the content-analysis systems so far developed are not adequate to the task of marking out the main variables in therapy. Most of them rely too much on the opinions presented by clients and neglect clients' unconscious motives. Some investigators have assumed that an increase in favorable opinions about himself indicates improvement in the client. While we believe that this is sometimes true, we are aware that clients can change their opinions in response to stimuli other than improved psychological well-being. For instance, a client having an unconscious motive to escape the anxiety evoked by therapy may offer the opinion that he feels much better and has been enormously helped. Therefore, the client says, he is ready to quit psychotherapy. Again, a client motivated by the desire to please the therapist may say that he has been greatly benefited. He would feel ungrateful if he did not acknowledge benefit. These two motives, used as illustrations, are only two of many possible motives, both conscious and unconscious, that can influence the client's expression of an opinion about himself. Our knowledge that such motivations exist warns us not to rely too much on these opinions.[5]

Content systems are inevitably criticized for what they leave out. The practicing clinician often feels that the measured part of the therapeutic transaction is pitifully small alongside the complex of stimuli that he senses as a participant observer. Yet it seems unfair to expect any single content-analysis system to describe all of this complex situation. We would probably make a fairer appraisal of content systems if we expected each system to deal with only a part of this complexity. An adequate descriptive and causal analysis of psychotherapy will most likely require a large number of measures, each of them shown to be reliable and valid for its limited purpose. Measures of the content of clients' and therapists' utterances will undoubtedly be supplemented by measures of other, nonverbal responses of client and therapist. By the combination of a variety of measures, each useful in its own domain, we may in time construct an adequate science of psychotherapy.

SUMMARY

Research on psychotherapy has been hampered until recently by the lack of permanent records of the transaction, by absence of objective measures, and by lack of an appropriate theoretical framework. The advent of sound recording of interviews, the widespread application of content-analysis methods, and the development of psychological theories such as learning theory have opened up new possibilities for research on psychotherapy. This paper surveys some of the first fruits of the new developments.

Studies are here classified as methodological (development of measures), descriptive, or theoretically oriented. An attempt has been made to include representative and important studies of each type. It is believed that the value of studies in this field must finally be assessed in terms of relevance to theory. Studies that only result in the development of new measuring instruments or in the publication of a description of some type of therapy are less valuable than studies that test some hypothesis. The methodological and descriptive investigations may be valuable, however, if they provide the tools and the hypotheses that make later theoretically guided investigation possible.

---

[5] We believe that an objective system of content analysis need not neglect unconscious motives. The therapist who attributes unconscious motives to his patient is, when he does this, reacting to cues provided by the patient's behavior. It should be possible to teach other persons to react similarly to the patient's behavior. If the responses of the patient that reveal unconscious motivation are verbal, it would be possible to designate these unconscious reactions by the content-analysis method. For example, if overconcern about another person's health indicates unconscious hostility toward that person, any sentence expressing overconcern can be scored as "unconscious hostility."

REFERENCES

1. Aronson, M. A study of the relationships between certain counselor and client characteristics in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 39–54.
2. Assum, A. L., & Levy, S. J. Analysis of a nondirective case with followup interview. *J. abnorm. soc. Psychol.*, 1948, **43**, 78–89.
3. Bales, R. F. *Interaction process analysis.* Cambridge: Addison-Wesley, 1950.
4. Berelson, B. *Content analysis in communication research.* Glencoe, Ill.: Free Press, 1952.
5. Bergman, D. Counseling method and client responses. *J. consult. Psychol.*, 1951, **15**, 216–224.
6. Blau, B. A. A comparison of more improved with less improved clients treated by client-centered methods. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 120–126.
7. Brody, E. B., Newman, R., & Redlich, F. C. Sound recording: the problem of evidence in psychiatry. *Science*, 1951, **113**, 379–380.
8. Bugental, J. F. T. A method for assessing self and not-self attitudes during the therapeutic series. *J. consult. Psychol.*, 1952, **16**, 435–439
9. Carnes, E. F., & Robinson, F. P. The role of client talk in the counseling interview. *Educ. psychol. Measmt*, 1948, **8**, 635–644.
10. Cofer, C. N., & Chance, June. The discomfort-relief quotient in published cases of counseling and psychotherapy. *J. Psychol.*, 1950, **29**, 219–224.
11. Coffey, H. S., Freedman, M., Leary, T., & Ossorio, A. Community service and social research—group psychotherapy in a church program. *J. soc. Issues*, 1950, **6**, 1–65.
12. Colby, K. M. *A primer for psychotherapists.* New York: Ronald Press, 1951.
13. Collier, R. M. A scale for rating the responses of the psychotherapist. *J. consult. Psychol.*, 1953, **17**, 321–326.
14. Covner, B. J. Studies in phonographic recordings of verbal material: I. The use of phonographic recordings in counseling practice and research. *J. consult. Psychol.*, 1942, **6**, 105–113.
15. Covner, B. J. Studies in phonographic recordings of verbal material: III. The completeness and accuracy of counseling interview reports. *J. gen. Psychol.*, 1944, **30**, 181–203.
16. Covner, B. J. Studies in phonographic recordings of verbal material: IV. Written reports of interviews. *J. appl. Psychol.*, 1944, **28**, 89–98.
17. Curran, C. A. *Personality factors in counseling.* New York: Grune & Stratton, 1945.
18. Dittmann, A. T. The interpersonal process in psychotherapy: development of a research method. *J. abnorm. soc. Psychol.*, 1952, **47**, 236–244.
19. Dollard, J., Auld, F., & White, Alice M. *Steps in psychotherapy.* New York: Macmillan, 1953.
20. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
21. Dollard, J., & Mowrer, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3–32.
22. Elton, C. F. A study of client responsibility: counselor technique or interview outcome? *Educ. psychol. Measmt*, 1950, **10**, 728–737.
23. Freud, S. Further recommendations in the technique of psychoanalysis. On beginning the treatment. The question of the first communications. The dynamics of the cure. In *Collected papers*, Vol. 2. London: Hogarth and the Institute of Psychoanalysis, 1924. Pp. 342–365.
24. Freud, S. Further recommendations in the technique of psychoanalysis. Recollection, repetition, and working through. In *Collected papers*, Vol. 2. London: Hogarth and the Institute of Psychoanalysis, 1924. Pp. 366–376.
25. Geller, A., Kaplan, D., & Lasswell, H. D. An experimental comparison of four ways of coding editorial content. *Journ. Quart.*, 1942, **19**, 362–370.
26. Gill, M., Newman, R., & Redlich, F. C. *The initial interview in psychiatric practice.* New York: International Universities Press, 1954.
27. Gillespie, J. F. Verbal signs of resistance in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 105–119.
28. Gorlow, L., Hoch, E. L., & Telschow, E. F. *The nature of nondirective group psychotherapy: an experimental investi-*

gation. New York: Teachers Coll., Columbia Univer., 1952.

29. Grummon, D. L. An investigation into the use of grammatical and psycho-grammatical categories of language for the study of personality and psychotherapy. Unpublished doctor's dissertation, Univer. of Chicago, 1950. (Not seen; cited by Mowrer [60].)
30. Gump, P. V. A statistical investigation of one psychoanalytic approach and a comparison of it with non-directive therapy. Unpublished master's thesis, Ohio State Univer., 1944.
31. Haigh, G. Defensive behavior in client-centered therapy. *J. consult. Psychol.*, 1949, **13**, 181–189.
32. Hilgard, E. R. Experimental approaches to psychoanalysis. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science.* Stanford, Calif.: Stanford Univer. Press, 1952. Pp. 3–45.
33. Hoffman, A. E. Reported behavior changes in counseling. *J. consult. Psychol.*, 1949, **13**, 190–195.
34. Hogan, R. A. A measure of client defensiveness. In W. Wolff and J. A. Precker (Eds.), *Success in psychotherapy.* New York: Grune & Stratton, 1952. Pp. 112–142.
35. Horwitz, Florence S. The relationship between a method of measuring tension and two methods of measuring the attitude toward self in thirty-six electrically recorded initial social casework interviews. M.S. thesis, Boston Univer. Sch. of Soc. Work, 1951.
36. Hunt, J. McV. Measuring the effects of social casework. *Trans. N. Y. Acad. Sci.*, 1947, **9**, 78–88.
37. Hunt, J. McV. Measuring movement in casework. *J. soc. Casewk*, 1948, **29**, 343–351.
38. Hunt, J. McV. A social agency as a setting for research—the Institute of Welfare Research. *J. consult. Psychol.*, 1949, **13**, 69–81.
39. Janis, I. L. The problems of validating content analysis. In H. D. Lasswell, N. Leites, *et al., Language of politics.* New York: Stewart, 1949. Pp. 55–82.
40. Jones, E. *The life and work of Sigmund Freud*, Vol. 1. New York: Basic Books, 1953.
41. Kahn, M. W. The role of perceptual consistency and generalization change in Rorschach and psychotherapy behavior. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953, Pp. 115–134.
42. Kaplan, A., & Goldsen, J. M. The reliability of content-analysis categories. In H. D. Lasswell, N. Leites, *et al., Language of politics.* New York: Stewart, 1949. Pp. 83–112.
43. Kauffman, P. E., & Raimy, V. C. Two methods of assessing therapeutic progress. *J. abnorm. soc. Psychol.*, 1949, **44**, 379–385.
44. Kubie, L. S. Problems in clinical research. *Amer. J. Orthopsychiat.*, 1947, **17**, 196–203.
45. Lasswell, H. D. Verbal references and physiological changes during the psychoanalytic interview: a preliminary communication. *Psychoanal. Rev.*, 1935, **22**, 10–24.
46. Lasswell, H. D. Certain prognostic changes during trial (psychoanalytic) interviews. *Psychoanal. Rev.*, 1936, **23**, 241–247.
47. Lasswell, H. D. Veraenderungen an einer Versuchsperson waehrend einer Kurzen Folge von psychoanalytischen Interviews. *Imago*, 1937, **23**, 375–380.
48. Lasswell, H. D. A provisional classification of symbol data. *Psychiatry*, 1938, **1**, 197–204.
49. Lasswell, H. D., Leites, N., *et al. Language of politics.* New York: Stewart, 1949.
50. Lewis, Virginia W. Changing the behavior of adolescent girls. *Arch. Psychol.*, 1943, No. 279.
51. Lindner, R. M. *Rebel without a cause.* New York: Grune & Stratton, 1944.
52. McIntyre, C. J. Acceptance by others and its relation to acceptance of self and others. *J. abnorm. soc. Psychol.*, 1952, **47**, 624–625.
53. Meadow, A., Greenblatt, M., Levine, J., & Solomon, H. C. The Discomfort-Relief Quotient as a measure of tension and adjustment. *J. abnorm. soc. Psychol.*, 1952, **47**, 658–661.
54. Miller, Helen E. "Acceptance" and related attitudes as demonstrated in psychotherapeutic interviews. *J. clin. Psychol.*, 1949, **5**, 83–87.
55. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald Press, 1944. Pp. 431–465.
56. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
57. Miller, N. E. Comments on theoretical models, illustrated by the development

of a theory of conflict behavior. *J. Pers.*, 1951, **20**, 82–100.

58. MILLER, N. E., & KRAELING, DORIS. Displacement: greater generalization of approach than avoidance in a generalized approach-avoidance conflict. *J. exp. Psychol.*, 1952, **43**, 217–221.

59. MILLER, N. E., & MURRAY, E. J. Displacement and conflict: learnable drive as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.*, 1952, **43**, 227–231.

60. MOWRER, O. H. Changes in verbal behavior during psychotherapy. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research.* New York: Ronald Press, 1953. Pp. 463–545.

61. MOWRER, O. H., HUNT, J. McV., & KOGAN, L. S. Further studies utilizing the Discomfort-Relief Quotient. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research.* New York: Ronald Press, 1953. Pp. 257–295.

62. MOWRER, O. H., LIGHT, B. H., LURIA, ZELLA, & ZELENY, MARJORIE P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research.* New York: Ronald Press, 1953. Pp. 546–640.

63. MURRAY, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.*, 1954, **49**, 305–310.

64. MURRAY, E. J. A method for studying psychotherapy. Unpublished doctor's dissertation, Yale Univer., 1955.

65. MURRAY, E. J., AULD, F., JR., & WHITE, ALICE M. A psychotherapy case showing progress but no decrease in the Discomfort-Relief Quotient. *J. consult. Psychol.*, 1954, **18**, 349–353.

66. MURRAY, E. J., & BERKUN, M. M. Displacement as a resolution of conflict. *J. abnorm. soc. Psychol.*, 1955, **1**, 47–56.

67. MURRAY, E. J., & MILLER, N. E. Displacement: steeper gradient of generalization of avoidance than of approach with age of habit controlled. *J. exp. Psychol.*, 1952, **43**, 222–226.

68. PAGE, H. A. An assessment of the predictive value of certain language measures in psychotherapeutic counseling. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 88–93.

69. PERRY, W. G., & ESTES, S. G. The collaboration of client and counselor. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research.* New York: Ronald Press, 1953. Pp. 95–119.

70. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. I. The development. *Educ. psychol. Measmt*, 1943, **3**, 105–126.

71. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. II. The evaluation. *Educ. psychol. Measmt*, 1943, **3**, 215–238.

72. PORTER, E. H. *An introduction to therapeutic counseling.* Boston: Houghton Mifflin, 1950.

73. RAIMY, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153–163.

74. RAKUSIN, J. M. The role of Rorschach variability in the prediction of client behavior during psychotherapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 60–74.

75. RASKIN, N. J. An objective study of the locus-of-evaluation factor in psychotherapy. In W. Wolff and J. A. Precker (Eds.), *Success in psychotherapy.* New York: Grune & Stratton, 1952. Pp. 143–162

76. REDLICH, F. C., DOLLARD, J., & NEWMAN, R. High fidelity recording of psychotherapeutic interviews. *Amer. J. Psychiat.*, 1950, **107**, 42–48.

77. REID, DOROTHY, & SNYDER, W. U. Experiment in "recognition of feeling" in non-directive psychotherapy. *J. clin. Psychol.*, 1947, **3**, 128–135.

78. ROBERTS, B. H., & STRODTBECK, F. L. Interaction process differences between groups of paranoid schizophrenic and depressed patients. *Int. J. group Psychother.*, 1953, **3**, 29–41.

79. ROBINSON, F. P. *Principles and procedures in student counseling.* New York: Harper, 1950.

80. ROGERS, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.

81. ROGERS, C. R. Electrically recorded interviews in improving psychotherapeutic techniques. *Amer. J. Orthopsychiat.*, 1942, **12**, 429–435.

82. ROGERS, C. R. The development of insight in a counseling relationship. *J. consult. Psychol.*, 1944, **8**, 331–341.

83. ROGERS, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.

84. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.

85. Rosenman, S. Changes in interpersonal relations in the course of non-directive counseling. Unpublished doctor's dissertation, Harvard Univer., 1953.

86. Roshal, Jean J. G. The type-token ratio as a measure of changes in behavior variability during psychotherapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 94–104.

87. Royer, Anne E. An analysis of counseling procedures in a non-directive approach. Unpublished master's thesis, Ohio State Univer., 1943. (Not seen; cited by Snyder [91].)

88. Seeman, J. The process of non-directive therapy. *J. consult. Psychol.*, 1949, **13**, 157–168.

89. Sheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.

90. Sherman, Dorothy. An analysis of the dynamic relationship between counselor techniques and outcomes in larger units of the interview situation. Unpublished doctor's dissertation, Ohio State Univer., 1945. (Not seen; cited by Robinson [79].)

91. Snyder, W. U. An investigation of the nature of non-directive psychotherapy. *J. gen. Psychol.*, 1945, **33**, 193–223.

92. Snyder, W. U. A comparison of one unsuccessful with four successful nondirectively counseled cases. *J. consult. Psychol.*, 1947, **11**, 38–42.

93. Stock, Dorothy. The self concept and feelings toward others. *J. consult. Psychol.*, 1949, **13**, 176–180.

94. Strupp, H. H. An objective comparison of Rogerian and psychoanalytic techniques. *J. consult. Psychol.*, 1955, **19**, 1–7.

95. Strupp, H. H. An objective study of certain psychotherapeutic operations; the effect of the therapist's personal analysis upon his verbal techniques. *Amer. Psychologist*, 1954, **9**, 479–480. (Abstract)

96. Tucker, J. E. Measuring client progress in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State Coll., 1953. Pp. 55–59.

97. White, R. K. *Black boy*: A value analysis. *J. abnorm. soc. Psychol.*, 1947, **42**, 440–461.

98. Wilson, E. B., Jr. *An introduction to scientific research.* New York: McGraw-Hill, 1952.

99. Zimmerman, W., & Langdon, J. A preliminary attempt to establish criteria for measuring progress in psychotherapy. Unpublished manuscript, Univer. of Illinois. (Not seen; cited by Mowrer [60].)

Received July 7, 1954.

Psychological Bulletin
Vol. 52, No. 5, 1955

EMPLOYEE ATTITUDES AND EMPLOYEE PERFORMANCE[1]

ARTHUR H. BRAYFIELD AND WALTER H. CROCKETT
Kansas State College

The systematic investigation of employee attitudes is a relatively recent development in American business and industry. Although Houser and his associates (**26**) pioneered in this field in the early 1920's there was little active interest until early in World War II when employee attitude surveys began to flourish (**49**, p. 7). Currently there is an abundant and growing literature on the use of this personnel tool (**56**).

Only infrequently, however, are discussions of the correlates of employee attitudes found and these are almost never substantiated by empirical evidence. Where we have located relevant discussions in the personnel and psychological literature a common assumption predominates—employee attitudes bear a significant relationship to employee performance. These are sample quotations: "... morale is not an abstraction. Rather it is concrete in the sense that it directly affects the quality and quantity of an individual's output." "Numerous investigations have established the certainty that productive efficiency fluctuates with variations in interest and morale." "... employee morale ... reduces turnover. It makes labor trouble and strikes less likely. It cuts down absenteeism and tardiness; lifts production."

It is of some practical and theoretical interest to establish the relationships which exist between employee attitudes and employee performance. The purpose of this review is to examine and summarize the empirical literature which bears upon these relationships and to engage in some discussion of the methodological and theoretical considerations involved in such investigations.

Examination of the literature reveals that it is (*a*) recent, and (*b*) frequently peripheral in the sense that relevant data were collected and analyzed incidental to some other objective.

We have established certain conditions for the inclusion of materials in this review. First, the indices of employee attitudes must permit classification of respondents along some attitude continuum. Second, the indices of employee attitudes must have been obtained directly from the employees themselves. Although we are willing to include ratings of job performance by supervisors and others, if no other criteria of performance are available, we are not willing to accept estimates of *attitudes* by someone other than the individuals themselves. Performances, we would contend, are less easily disguised by the individual and less readily distorted by the observer than are attitudes. Third, the investigations must have been conducted in industrial or occupational settings. Within the limitations of interlibrary loan service our coverage is complete through July, 1954. We have made no effort to unearth unpublished studies although we report several including three studies by one of us.

---

[1] Portions of this paper were presented by Brayfield before the University of Minnesota Chapter of Psi Chi in March, 1954.

The following scheme was adopted as a convenient and meaningful way of categorizing the literature.

1. Daniel Katz and Robert Kahn (**33**, p. 657) have suggested that "in social structures it is important to distinguish between: (1) the motivation to stay within the system, to remain a part of the group and (2) the motivation to act in a differential manner within that system." We have thus distinguished between those studies which involve performance on the job and those which involve withdrawal from the job (absences, accidents, turnover).

2. Within the above breakdown we have made a further differentiation based upon research design. One major design relates the attitudes of individuals to their performances as individuals. A second design relates the attitudes of the members of groups to their performances as groups.

3. A still further classification differentiates between studies in which a single index of attitudes either as a single item or as a summation of items was used and those few in which multiple indices were used.

We have not attempted to define such terms as job satisfaction or morale. Instead, we have found it necessary to assume that the measuring operations define the variables involved. Definitions are conspicuous by their absence in most current work in this area.

Where reliability data are reported for the attitude and performance measures, we have included them in our summaries. We also have attempted to specify whether or not the attitude data were collected under conditions which preserved the anonymity of the subjects. Throughout the first section of the paper we have tried to hold comments on methodology to a minimum, postponing detailed methodological considerations until the substantive material has been covered.

Before summarizing and discussing the literature it may be appropriate to describe the investigation which, as far as we can determine, initiated research in this area of industrial psychology. The classic study relating attitudes and performance in an industrial setting was conducted by Kornhauser and Sharp (**39**) in 1930 in Neenah, Wisconsin, in the mill operated by the Kimberly-Clark Corporation. Between 200 and 300 young girls engaged in routine repetitive jobs at machines were studied. Both questionnaires and interviews were used. The questionnaires were patterned after those developed by Houser and covered a range of specific attitudes—toward supervisors, repetitiveness and speed of work, personnel policies, wages, and the like. Scores were computed for groups of items and item responses were analyzed. Intercorrelations among different item groups ran about .4 to .5. Reliabilities were thought to be somewhat higher.

The finding on relationship of attitudes to performance is summed up in the statement that "Efficiency ratings of employees showed no relationship to their attitudes." No description is given of the rating system. Further, the authors say, "In one group of 20 girls for whom we had comparable output records, three of the four with the most unfavorable attitudes were first, second, and fourth in production and the two most favorable were near the bottom in production."

With respect to the criterion of withdrawal from the job, Kornhauser and Sharp reported that "Unfavorableness of job attitudes is slightly correlated with lost time because of sickness."

Relations between attitudes and intelligence, age, schooling, marital status, home life, emotional adjustment, and supervision also were studied. This early report should be read by anyone seriously interested in this area of investigation.

PERFORMANCE ON THE JOB

Individual Analysis

Three unpublished studies have used the Brayfield-Rothe Job Satisfaction Blank as an index of job satisfaction. In 1943 Brayfield (4) started work on the development of a scale intended to give what might be called a global measure of job satisfaction. It was predicated on attitude theory and applied the Thurstone scaling technique. After some preliminary work Likert's scoring technique was applied to 18 Thurstone-scaled items to produce an index which had a range of scores from 18 through 90 with a neutral or indifferent point at 54. The resulting scale gave a corrected split-half reliability coefficient of .87 when used with 231 women office employees. It differentiated between adults enrolled in a night class in personnel psychology who were employed in personnel jobs and similar students who were employed in non-personnel jobs. For the same group, a correlation of .91 was obtained between the Brayfield-Rothe and the Hoppock Job Satisfaction Blank.

In 1944, in connection with a larger study Brayfield collected data on 231 women office employees working for the same firm but employed in 22 different offices throughout the country. The scale was administered to small groups of individuals as part of a test battery. All materials were signed. At the same time supervisor's ratings on a graphic rating scale were obtained for all employees in the sample. A total score was computed from three items covering quantity, quality, and over-all worth to the company. About two-thirds of the employees were rated by two supervisors and the ratings were averaged. Reliabilities of ratings are unknown although in one office it was possible to compare two supervisors who had rated the same 23 women. The intercorrelation was in the low seventies.

When job satisfaction scores for these women clerical workers were compared with their performance ratings a correlation of .138, significant at the 5% level, was found. To control for the influence of job level, the 231 women were classified into six groups as follows: Stenographers (50); General Clerical (40); Typists (38); High Level Machine Clerical (36); Low Level Machine Clerical (34); Entry (33). The correlations for the first five groups ranged from −.06 to +.13. None were significant. The correlation for the group of 33 inexperienced and untrained girls (Entry) was .387, significant at the 5% level. An additional group of 35 women telephone order clerks provided a correlation of .26 which was not significant.

In 1950, Brayfield and Mangelsdorf obtained data on 55 second-, third-, and fourth-year plumber apprentices employed in a number of firms in Oakland, California. All were enrolled four hours per week in a public vocational school. The subjects completed the Brayfield-Rothe job satisfaction scale during classes as part of a testing program in which all the materials were identified by name of respondent. The corrected split-half reliability coefficient was .83. Performance ratings were obtained for each plumber from his foreman or employer. The rating form consisted of 25 scaled items in check list form developed by Goertzel (**19**, p. 117) who attempted to provide a generalized scale that could be used for assessment of workers on any type of job. For various groups of workers Goertzel found a correlation of approximately .80 between ratings on two forms of 25 items each. The correlation between job satisfaction scores and ratings was .203 which is not significant.

In 1953, Brayfield and Marsh studied the measured characteristics of 50 farmers enrolled four hours per week in a veterans' on-job training program. The median age of the subjects was in the early thirties. They had lived on farms most of their lives; all were managing their own farms. Among other materials they completed the Brayfield-Rothe job satisfication scale. All materials were signed. The corrected split-half reliability coefficient was .60; if the three subjects with the most inconsistent responses had been eliminated, the reliability coefficient would have become .77.

The subjects' performance as farmers was rated by their instructors who ranked them in order of effectiveness. Sixteen farmers were ranked by one instructor, 14 by a second, and the remaining 20 by another. Ranks were transmuted into "scores" (18) and cast into a single distribution. Rerankings after several months, when similarly treated, correlated .86 with the original rankings. Instructors were not aware that they would be asked to re-rank their students.

For the 50 farmers the correlation between job satisfaction scores and performance ratings was .115 which is not significant. If the three "erratic" subjects had been eliminated, the correlation would have become .133.

The same job satisfaction scale was used in 1953 in an unpublished study by Roger Bellows and associates of 109 Air Force control tower operators.[2] The correlation with individual proficiency ratings was .005.

Gadel and Kriedt (17) report a study employing a design similar to that used in the investigations just described. One hundred and ninety-three male IBM operators working in the machine rooms of numerous divisions of the Prudential Insurance Company home office completed and signed a 10-item job satisfaction questionnaire "designed to cover a variety of attitudes related to work duties." The performance criterion consisted of rank-order ratings on over-all job performance made by the immediate supervisor. Ratings were converted to standard scores and correlations were computed for each of the groups. The resulting correlations were averaged using the Fisher z transformation. The relationship between job satisfaction and performance was found to be .08.

The Life Insurance Agency Management Association has engaged in job satisfaction studies since the early 1940's. A report of their work which falls into the research classification under consideration was published by Habbe (23) in 1947. Job satisfaction questionnaires were mailed out to 9,353 insurance agents. Seventy-five per cent were returned of which more than 90% were usable. Signatures were not requested although quite a few agents did identify themselves. The blank contained questions asking about single phases of the job to be answered by one of five alternatives indicating satisfaction or dissatisfaction. A single question asked "How do you feel about your job as a life underwriter?" The performance rating was in the form of a self-report since each agent was asked to check whether his previous year's production was "under

---

[2] R. M. Bellows. Personal communication. June 30, 1954.

$200,000" of insurance or $200,000 or over." Agents producing under $200,000 scored 4.15 on what evidently was the single general satisfaction item as compared to 4.11 for the high producers. The "Extremely Satisfied" score is 5.00. The relationship is insignificant or slightly in favor of the lower producers. It should be noted that this performance criterion is a self-report and that the break at the $200,000 point might not be the best point for analyzing the relationship.

Baxter and his associates (1) have recently reported a training evaluation study concerned with new debit insurance agents (service and sell weekly and monthly premium, ordinary, and group insurance for families within a specific geographical territory). Included in the data collected were responses to a comprehensive job satisfaction attitude questionnaire with items varying in number from 32 to 43 depending upon when it was administered. Respondents apparently were identified. Suvervisor ratings on a 5-point, 9-item graphic rating scale were collected. Sales volume figures were obtained for each agent for his first year on the job. Although the correlations between the job satisfaction index and the performance criteria were not reported, the investigators have made them available.[3] For 223 agents the correlation between satisfaction and supervisor's rating is .23, significant at the 1% level. The correlation between satisfaction and sales volume is .26 also significant at the 1% level. This is the only study in this classification which uses an objective performance criterion. The incentive situation is also more clear-cut here except perhaps for the farmers. Although this correlation is significant, it is quite low.

One of the most carefully done studies which we have inspected is Mossin's (48) investigation of the selling performance and what he termed contentment of 94 teen-age female retail sales clerks in a large New York department store. His performance criteria of 12 items were based on the ratings of four experienced and specially trained shoppers. Ratings on five items formed a composite labeled "Selling Attitudes." Ratings on three other items were combined as an index of "Selling Skills." The intercorrelation was .76. In addition, the entire 12 items formed a composite which correlated beyond .9 with each of the other criteria. Several detailed analyses of the reliability of the criteria were made including intercorrelations among the four shoppers. A minimum estimate of the reliability of the criteria would be that they exceeded .7 and might actually have been somewhat higher.

Mossin used two job satisfaction measures. One was an over-all composite rating secured by combining the scores on 6 attitude items inquiring about "affective dispositions" toward departmental assignment, merchandise assignment, relations with customers, relations with fellow salesgirls, relations with supervisors, and working conditions, along with one item regarding intention to remain in retail selling plus one item requiring a self-appraisal of sales ability. The second index was a single multiple-response item asking "How you REALLY feel toward your job." Responses on these two indices of job satisfaction were obtained during an individual data collection session with the investigator. Therefore, the respondents were identified. The correlation between the two satisfaction criteria was .53.

---

[3] B. Baxter. Personal communication, February 17, 1954.

The composite job satisfaction score correlated −.07 with the Attitudes criterion and −.03 with the Skill criterion. The single item job satisfaction index correlated .15 and .06, respectively. None of these is significant. No results were reported for the 12-item composite performance criterion although it may be inferred that they would be of approximately the same magnitude since it was highly correlated with the other two criteria. This is a carefully executed investigation and should be consulted by anyone working in this general area.

The final major investigation in this series, by Bernberg (2), is the only one to use differentiated attitude measures. He included a measure of group morale as identified by 34 "indirect method" items, a 12-item scale presumed to measure an employee's acceptance of the formal organization (e. g., "I think this company treats its employees worse than any other company does"), a 0–100 thermometer scale with seven verbal referent points based on the single statement, "On the whole, I believe that the supervisor in my group is a man who knows his job and is a leader," and a similar thermometer scale for the self-rating statement, "On the whole, I believe that my group has a high degree of morale. By that, I mean the men work willingly and cheerfully as a well organized team." The intercorrelations among the four indices as computed for 890 hourly paid workers in a large aircraft manufacturing plant ranged from .47 to .77 with the median at .5. Split-half reliabilities for the two multi-item scales were approximately .8.

Questionnaires embodying these measures were sent home with more than 1,000 employees of an aircraft plant. No returns were accepted after 48 hours by which time 88% were back. Presumably the respondents were identified. The performance criterion was the average weighted score of a graphic rating scale with the five dimensions of adaptability, dependability, job knowledge, quality, quantity. The split-half reliability was .8.

The correlations between the four attitude measures and the performance criterion ranged from .02 to .05.

Four miscellaneous studies warrant brief mention only. An English doctoral dissertation (40) is reported to include the finding that it was clearly determined that "there is almost no relationship between proficiency and satisfaction among (British) post office counter clerks."[4] Kerr (38) reports a master's study finding of a correlation of −.76 between a 10-item job satisfaction measure and employer reports on the frequency of what he termed grievance, advice, and catharsis conferences with employees in two very small Indiana plants. The study is relevant mainly as suggesting a possible performance criterion for investigation. Chase (7) has a very inadequately described study which purports to find a small positive relationship between superintendent's ratings and teacher's satisfaction. Brody's (5) master's thesis at New York University describes an investigation in which the relationship between Hoppock Job Satisfaction Blank scores and production under a piece work incentive plan correlated .68 for 40 employees. This is an extraordinary finding. However, examination of the raw data in the Appendix casts serious doubt on the meaningfulness of the correlation. Two groups working under different incentive conditions are lumped together. For the 22 cases which might actually be legitimate subjects, the Hoppock scores do not conform to any known appropriate scoring system for that particular Blank. The production data are bimodal.

---

[4] Reported by Heron, A. Industrial psychology. *Ann. Rev. Psychol.*, 1954, 5, 203–228.

At this point we can summarize the findings for this research design. The prototype study used a single over-all index of employee attitudes variously titled job satisfaction or morale. Respondents were identified. A distribution of individual scores was related to some index of individual performance on the job. Customarily, a single occupational group was studied. When 14 homogeneous occupational groups and one large sample of assorted hourly factory workers were studied, statistically significant low positive relationships between job satisfaction and job performance were found in two of the 15 comparisons. These results, pointing to an absence of relationships, are in line with the findings of the pioneering Kornhauser and Sharp investigation.

Group Analysis

The essentials of this design are as follows. Employee attitudes are determined individually but the average for the group or the percentage responding in a certain manner is related to some estimate of performance or productivity for the group as a whole. This arrangement requires at least two groups. Characteristically, comparisons are by departments within a firm rather than by occupation.

The antecedents of this approach are to be found in a study by Rensis Likert which was reported in a privately circulated document in 1941. We have not examined the report. According to a reference to it in one of Katz's (**30**) papers, the morale of insurance agents in 10 agencies rated superior in operational efficiency by the home offices of nine companies was compared to that of agents in 20 agencies rated below average. We infer that some form of attitude questionnaire was used since Likert conducted the study although interviews may have been involved. Katz says that "Morale was found to be significantly related to the criterion." This study is mainly of historical interest.

Three studies employing this design or a modification of it utilized a single index of employee attitudes. Katz and Hyman (**32**) report a study which they supervised during World War II under the general direction of Likert. Their concern was with employee morale in shipyards and its relation to productivity, among other considerations. Two summary measures of morale were used, both of which were obtained from personal interview protocols. One was a yes-no answer to the question: "Have you ever felt like quitting the yards?" The rank order of the percentage who had felt like quitting was compared with an index of productivity (time to turn out a ship) for the five shipyards being studied. The two rank orders agreed fairly well. The second measure of employee attitudes was furnished by the responses to 7 items regarding specific aspects of the working situation and environment. The relationship to productivity was somewhat less marked than the first comparison although the authors comment that "In general the yards with high productivity were the yards with high worker morale." It should be remarked that, although the productivity differences were very great among the yards, the morale differences were really quite small; the morale scores for the five yards were 9.3, 9.4, 10.0, 10.0, 10.9.

Giese and Ruter (**20**) employed this design in a study of employees of a small national mail order company. It is one of the few studies which had as its primary purpose the determination of the relationship in which we are interested. In fact, the aim of the study was to devise a method for predicting the morale of departments from objective data. The only description of the attitude measure is "A morale questionnaire was scored so that a quantitative score was available." There is no statement regarding the anonymity of the subjects. Three objective measures of efficiency were available. For each of 25 departments there were available three average measures of efficiency and one average morale score. When correlations based on group averages were computed they were found to range between .15 and .27. None of these is significant (our determination).

The Triple Audit studies (**62**) of the Industrial Relations Center at the University of Minnesota fall into this research design category. Here the firm was the basic unit of comparison. Since the number of firms studied was small, only seven, the authors advise that it is impossible to draw any conclusions about the relationships obtained.

The next series of studies reviewed here used the same design but differ from the three just mentioned in that they make some differentiation among employee attitudes. That is, they make some attempt to specify component parts.

The early work of Likert and Katz has been continued by them at the University of Michigan since the war. The prototype study was undertaken in 1947 in the Prudential Insurance Company and findings were reported in some detail in 1950 (**34**). Although the objectives were broad, an important portion of the study was devoted to exploring the relationships between employee attitudes and productivity.

One and one-half hour free answer interviews covering 53 questions were held with 419 nonsupervisory clerical workers in 24 different sections. Responses were coded. The sections were arranged in parallel pairs in order to hold constant as many factors as possible. One set of 12 sections was designated as high productivity on the basis of production records while the parallel set was composed of low productivity sections. The authors note that productivity differences between the pairs were not great, rarely more than 10 per cent. Each of the high-low productivity pairs consisted of two sections handling the same type of work with the same type of people at the same job levels and were very similar on a number of factors.

A unique feature was the construction of four indices of attitudinal variables. The differentiations were made on a theoretical basis with some empirical confirmation for the relationship among the items used in each index. Four variables were specified: (*a*) pride in work group; (*b*) intrinsic job satisfaction; (*c*) company involvement; and (*d*) financial and job status satisfaction. Pride in work group was the most independent of the four; the remaining three intercorrelated around .4.

When these morale indices were related to productivity, only pride in work group showed a distinct relationship. Productivity groups were also differentiated by three specific attitude items not included in the morale indices.

A second study of similar design investigated these relationships among section hand employees of the Chesapeake & Ohio Railroad (**35**). Somewhat different morale items with more emphasis upon individual items were used in intensive interviews. Productivity criteria consisted of over-all quality and quantity ratings by supervisors. There was some slight support for the previous finding of a relationship between pride in work group and productivity. The authors emphasize the lack of relationships found between employee attitudes and productivity.

Both Michigan studies, but particularly the insurance company one, may be studied with profit. The investigators are self-critical and also provide hypotheses for further investigation. The investigations are excellent examples of both the virtues and the shortcomings of survey techniques. Both illustrate attempts to measure a large number of variables with less precision perhaps than is ultimately desirable. These well-publicized studies are important in our present context because they have called into question a common assumption about an important relationship, have perhaps stimulated research elsewhere, and have produced a great amount of theorizing about motivation in industry.

Three recent investigations are patterned somewhat after the Michigan inquiries. Two studies by Comrey and associates (**9**, **10**) are considerably less well reported. The findings, among Forest Service and Employment Service personnel, lend some slight support to the Michigan report of a relationship between attitudes toward the group and performance. Weschler (**60**) found a slightly negative relationship between a single-item index of job satisfaction and production among employees in two comparable groups of a Naval research laboratory. He obtained a similar result for a single item index of work group morale.

A recent study by Lawshe and Nagle (**41**) warrants extended comment. Two hundred and eight nonsupervisory office employees in 14 work groups at a plant of the International Harvester Company completed a 22-item questionnaire which was described by the authors as an attitude toward supervisor measure. The corrected split-half reliability was .92. There is no report regarding the anonymity of the respondents. The scores were related to group productivity.

A paired comparison rating of productivity by six plant executives was used. Each executive compared from 8 to 14 work groups under instructions to indicate " . . . The department in each pair which is, in your opinion, doing its job better." Ratings were converted to standard scores and averaged. The reliability of the means of all six raters was estimated to be .88. The authors are careful to point out that "one does not know for sure what the raters really had in mind when they rated." They suggest that "How little trouble the work group caused, whether or not it had the answers when called upon, whether or not it could cope with rush situations, and similar considerations are believed to have been the prime factors in the executives' ratings."

The average rating of each work group was correlated with average attitude toward the supervisor score in the work group. The resulting Pearson coefficient was .86, significant at the 1% level for $N=14$. This is, of course, a remarkable result. A whole superstructure of industrial psychology could well be erected on this finding; stranger things have happened. However, the authors sound a note of caution: "On the basis of this study it can be concluded only that the behavior of the supervisor, as perceived by the employees, is highly related to the productivity of the group as perceived by higher management."

It occurs to us that it may be a misnomer to call the questionnaire an attitude questionnaire. It might well be considered to be a supervisor behavior- or performance-rating device. For example, the questions included such things as, does he: give you straight answers, avoid you when he knows you want to see him about a problem, criticize you for happenings over which you have no control, delay in taking care of your complaints, keep you informed, give you recognition, show interest in your ideas, follow through on his promises, explain to you the "why" of an error to prevent recurrence, give you sufficient explanation of why a work change is necessary. There is supporting evidence for a finding that supervisory performance is related to productivity (**34**). It might be suggested further that the obtained correlation really expresses the relationship between supervisor performance ratings by employees and supervisor performance ratings by the executives since the performance of a work group may be judged in part at least on the basis of the observation of the supervisor and certainly on his reports. We suggest that the finding is relevant to current work on supervisor performance and productivity but we are skeptical of its direct relationship to the area of research being examined in this review despite the title.

The results from the study design which we have described in this section are substantially in agreement with the previous findings of minimal or no relationship between employee attitudes and performance. They do supply the hint that morale, as a group phenomenon, may bear a positive relationship to performance on the job.

WITHDRAWAL FROM THE JOB

As indicated earlier we have differentiated between performance on the job and withdrawal from the job. In this section we briefly summarize the trend of the evidence when employee attitudes are related to some form of withdrawal from the job. Withdrawal is indicated by absence and tardiness, by accidents (under one assumption), and by turnover or employment stability.

Absences

Individual analysis. The individual analysis design has been used in four studies. In another of the Triple Audit studies, Yoder and associates (**62**) employed a 66-item employee attitude questionnaire which yielded a total score as an index of general attitude. Respondents apparently were unidentified. Absence and tardiness data were *furnished by the respondents* on the questionnaire face sheet. Five groups of employees were studied including office workers, department store personnel, and manufacturing employees. No statistically significant relationships were found between the attitude index and absences. One significant relationship was found for tardiness. Four others were insignificant.

In a study of worker attitudes toward merit rating, Van Zelst and Kerr (**55**) include data relevant to our topic. Three hundred and forty employees selected by their employers in 14 firms out of the 50 invited to participate furnished a *self-report* of their absences and tardinesses. Two Hoppock-type job satisfaction items were combined to give a single index. Respondents apparently were anonymous. Job satisfaction correlated .31 with a favorable absentee record and .26 with a favorable tardiness record. These are significant at the 1% level.

Bernberg (**2**) used four different measures of absence and tardiness which had split-half reliabilities in the seventies. These data were taken from the company records. The intercorrelations with the four measures of employee attitudes of the 890 aircraft plant workers ranged from −.05 to +.07.

Group analysis. The group analysis design also has been used in four studies. Giese and Ruter (**20**) found an insignificant relation between tardiness and a single morale score when the group averages of employees in 25 departments of a mail order house were correlated. However, the correlation of −.47 between the morale index and absences is significant at the 5% level (our determination).

Kerr and associates (**38**) obtained mean scores on his 10-item Tear Ballot job satisfaction blank for the employees in 30 departments of a Chicago plant presumably under conditions of anonymity. They used six measures of absenteeism. This is a major contribution of the study since there are numerous problems in indexing absenteeism. The importance of such an analytical approach is evident when it is observed that job satisfaction correlated .51 with total absenteeism rate but correlated −.44 with unexcused absenteeism. One other relationship was statistically significant.

In their wartime morale studies, Katz and Hyman (**31**) used six specific attitude items. These morale indices were positively related to absenteeism. The magnitude of the relationship is typified by responses to this item: 44% of the workers disliking their jobs were categorized as absentees as compared with 36% who liked their jobs.

Perhaps the most extensive investigation has been made by Metzner and Mann (**46**) of the Michigan Survey Research Center. The data were collected in the Detroit Edison Company according to a design similar to the Prudential and Chesapeake & Ohio studies. Anonymous questionnaires provided the attitudinal data. White-collar and blue-collar men and white-collar women were the subjects. The most striking finding was that there was no relationship between absences and attitudes toward any aspect of the work situation for white-collar women. Among white-collar men, 10 out of 15 attitudinal measures showed significant relationships at the 10% level. Eight of these were significant at the 5% or 1% levels. However, when job level or grade was controlled to some extent by grouping into high- and low-skill levels, there was practically no relationship between attitudes and absences for the high-skill level jobs for the seven items it was possible to study. A fairly consistent relationship remained for the low-skill level jobs. Among the 18 items used with the 251 blue-collar men, nine were significant at the 10% level or better, six being at the 5% level or better. Incidentally, these are all percentage differences among various absence categories and the adjacent category differences are not particularly impressive although the differences between the extreme categories are appreciable.

ACCIDENTS

Hill and Trist (**25**), English investigators, have recently suggested that "Accidents (may) be considered as a means of withdrawal from the work situation through which the individual may take up the role of absentee in a way acceptable both to himself and to his employing organization." Accidents are considered to involve the "quality of the relationship obtaining between employees and their place of work." In an empirical test of this hypothesis they found accident rates to be positively associated with other forms of absences and to be most strongly associated with the least sanctioned forms of absence. Their study does not, of course, bear directly on our immediate concern. We include it to indicate a possible linkage with absence data.

Group analysis. We have found two studies on the relationship between employee attitudes and accidents. Stagner and associates (**53**) used a group analysis design to study the job satisfaction of railroad employees. A total of 715 employees in 10 divisional groups, 2 accounting offices, and 12 shops were included. Fifteen specific items of an apparently anonymously administered questionnaire were given arbitrary weights based on the percentage of employees checking and were summed to give a single job satisfaction index. Mean satisfaction scores by groups were correlated with group accident rates. The obtained correlations are negative and small. Surprisingly, the authors conclude that "We thus feel considerable confidence in the conclusion that working in a group with a high accident rate will tend to make the individual worker anxious, and reduce his satisfaction with his job." However, the correlations are not statistically significant (our determination) and the causal sequence indicated in the quote is speculation.

The Triple Audit (**62**) studies also considered accidents. Although the authors have entered a general disclaimer as to the significance of their group design findings, the accident finding is intriguing. Employees in the three firms with fewer than average accidents had a mean attitude score of 133 while the employees in the three firms with more than average accidents had a mean score of 143. From these limited data it appears that there is a tendency for the firms with higher accident rates to have more favorable attitude scores. The data are interesting but should not be given disproportionate emphasis.

EMPLOYMENT STABILITY

Individual analysis. Employment stability remains to be studied. In a study comparing indirect and direct methods of appraising employee attitudes, Weitz and Nuckols (**59**) provide relevant although peripheral data. This is an individual analysis design. Two attitude questionnaires, one composed of 18 indirect items and one consisting of 10 direct questions, were mailed to more than 1,200 insurance agents representing one company in the southern states. Forty-seven per cent submitted answers. The respondents were identified. Total scores for each of the questionnaires were then related to survival during a one-year period. The direct method correlated .20, significant at the 1% level; the indirect method correlated insignificantly with survival. There was some sample bias resulting from the fact that a disproportionately small number of men, who subsequently terminated, responded.

Kerr (**37**) correlated total Tear Ballot job satisfaction scores obtained individually but without identification from 98 miscellaneous wage earners with an index of *self-reported* past job tenure (number of years on labor market divided by number of employers). The result for an unweighted total satisfaction score was .25, significant at the 5% level. Thus there seems to be a slight positive relationship between attitude toward present employment and past employment stability among the members of a heterogeneous group making self-reports on their employment records.

Van Zelst and Kerr (**55**) correlated total score on two Hoppock-type items with *employee reports* of previous job tenure. The obtained correlation was .09.

Friesen (**16**) recently has attempted to measure employee attitudes using an incomplete sentences technique. He developed four scales comprising a total of 81 items. These he labeled Working Situation, Work, Self, Leisure. He studied women office workers from one company with *N*'s ranging from 38 to 70. The blanks were signed. Split-half reliability coefficients for the four scales ranged from .68 to .82. Intercorrelations among the scales ranged from .26 to .72 with Working Situation and Work being the two most highly correlated. Friesen attempted to validate the scales by obtaining modified "Guess Who" ratings from seven to nine fellow employees for each member of his sample. These ratings had uncorrected reliability coefficients ranging from .57 to .78. Their obtained correlations with the attitude scales were moderately high being .59, .67, .45, and .52, respectively. When the four attitude scales were related to a criterion of employment stability (two or more years with each employer versus less than two years with each employer) the biserial correlations were .43, .53, .37, and .22, respectively. All were significant at the 4% level or better. This is a *retrospective* measure of employment stability.

Two attitude items, chance to make decisions on the job, and a feeling they were making or had made an important contribution to the success of the company, were significantly related to turnover in a study by Wickert (**61**). This study has the limitation that the employees who had left the company were interviewed after their departure.

Group analysis. The group analysis design has been used in three studies previously described. In the Giese and Ruter (**20**) investigation, morale scores correlated −.42 with a per cent turnover criterion for 25 departments. This is significant at the 5% level (our determination). Kerr (**38**) found the relationship between total Tear Ballot score and turnover in 30 departments to be −.13 which is not significant. The Triple Audit studies (**62**) found average monthly turnover to be unrelated to attitudes in seven companies.

With respect to withdrawal from the job, then, there is some evidence, mainly from the group design studies, of a significant but complex relationship between employee attitudes and absences. The investigations reviewed here also lend some support to the assumption that employee attitudes and employment stability are positively related. The data on accidents and attitudes are extremely limited, but they do not support any significant relationships.

In summary, it appears that there is little evidence in the available literature that employee attitudes of the type usually measured in morale surveys bear any simple—or, for that matter, appreciable—relationship to performance on the job. The data are suggestive mainly of a relationship between attitudes and two forms of withdrawal from the job. This tentative conclusion, contrary as it is to rather widely held beliefs, warrants an attempt to identify and evaluate some of the factors which may account for these results.

We have chosen to comment on

some methodological considerations, followed by a discussion of theoretical issues, and concluding with some possible implications for future research.

METHODOLOGICAL CONSIDERATIONS

The methodological questions which might be raised about any field study are legion (**15**, **27**, **28**, **36**, **44**). We shall comment briefly on some methodological limitations of the studies we have reviewed and on some fairly general problems of analysis and design of field study which are relevant to our topic.

Limitations of the Current Literature

It is difficult to know whether inadequacies in research reports reflect faults in research design, or whether space limitations in journals force authors to omit a considerable part of the methodological detail of their studies. This problem is apparent in the following discussion of three methodological areas: sampling, measurements, and the general procedure of the study.

Sampling. In most industrial studies there is sampling of both respondents and items. With regard to the sampling of respondents, reports frequently fail to state how respondents were selected, the possible selective biases, or the population which the sample is supposed to represent. It is possible that conflicting results from two studies may reflect differences in the characteristics of the populations sampled. This would be of considerable interest theoretically and practically, but it is unlikely to be detected unless the respective populations are described in some detail.

With regard to sampling of items, it is not uncommon for questionnaires administered to industrial workers to contain 150 or more items, and for research reports to mention results from only a few of these items. One is left in doubt as to whether the items reported upon were selected on the basis of some theoretical analysis, in which case the other items presumably were not theoretically relevant to the issue in question, or whether the items were selected on the basis of statistical significance and the theoretical orientation developed in retrospect. The latter procedure obviously capitalizes on chance, and the results should be subjected to further research before being accepted with any great confidence. Unless researchers inform the reader when they have operated on an *ad hoc* basis the tentative nature of the results will not be apparent.

Criterion measures. Research reports sometimes fail to describe the specific measurements that were used! In addition, there is extreme diversity in the kinds of employee attitudes that are measured, and in the questionnaires and interview schedules that are used to identify or measure them. This means that disparity in results between two or more studies may sometimes reflect differences in operational definitions, rather than differences between the populations being studied. We shall deal in greater detail with the measurement of attitudes in a subsequent section. Here we shall concentrate for the moment on criteria of job performance and of withdrawal from the job.

It is not our purpose to present a definitive report on problems involved in the selection and measurement of criteria. This area is well covered in current books in the field (**19**, ch. 3; **54**, ch. 5). However, it is obviously impossible to assess the relation between employee attitudes and job performance without some measure of the latter. Therefore, it is appropriate to dwell briefly on the criterion problem.

Let us consider, first, criteria of job performance. On production-line jobs as well as in sales and many other white-collar jobs, the criterion which seems, at first glance, best to reflect adequacy of job performance is productivity. Yet the measurement of productivity is much less simple than it may seem initially.

To be valid, a comparison of output between two individuals must equate the conditions under which the individuals operate. One salesman's territory may be potentially more fruitful than another's and his higher sales may cause, rather than result from, his greater satisfaction with his job. Two machines may vary in their potential output, in their state of repair, and in many other ways. Frequently there are, in addition, external restraints on productivity. Output in a factory may be determined by the speed of the assembly line or the speed of the machine, by the amount of material provided to an individual by some feeder line, or by the quality of the material being processed. Variation in situational factors such as these will affect total productivity no matter what the level of individual job performance.

Furthermore, on many jobs, white-collar as well as production-line, it is impossible to get a measure of productivity because a certain amount of work is required during the day and no more is produced, because adequate records are not kept, because the product depends on group rather than individual performance, or for a variety of other reasons.

Where output data are questionable or unobtainable as a performance criterion, it is sometimes suggested that other objective criteria such as quality of production as measured by number of errors, amount of scrap, and the like, be used. However, many of the above considerations such as machine variation, availability of records, and comparability of materials processed also apply to quality data.

Relevance, reliability, freedom from contamination, and practicality are criterion requisites which are not easily satisfied even when objective performance indices are available.

In the absence of direct counting measures of job performance, researchers are likely to obtain subjective evaluations, usually by superiors in the organization, of the performance levels of individuals or groups. Such ratings frequently were used in the investigations we have reviewed. The limitations and the precautions necessary in the use of ratings are well known and adequately documented (**19**, ch. 4; **22**, ch. 11). Problems such as the selection of factors to be rated, the errors of human judgment-making, contamination, and reliability plague the investigator who is forced to rely upon ratings.

Criterion selection is a problem also in studies attempting to predict absenteeism and turnover. Sex differences, differences of position in the organization, and similar factors influence absences in such a way that differences in absence rate between two groups may reflect many things other than differences in employee attitudes. The same thing may be said of turnover. General employment and wage levels, selective service policies, and similar factors will all affect the rate of turnover and may mask the effect of employee attitudes on turnover.

The selection of criteria, then, involves a choice among a number of possible measurements all of which may be affected by situational factors over which the investigator has little if any control.

Procedural problems. A frequent drawback of reports of the type summarized in this paper is a failure to report reliability data for the measurements, both of attitudes and of criteria of performance.

For the most part, the validity of the measures employed is unreported. With respect to the criterion measures this takes the form of failure to discuss the relevance of the particular criteria used. Usually employee attitude indices are assumed to have some form of face validity; empirical validation is seldom attempted. The reader is expected to assume that the questionnaires measured what they were intended to measure. In view of the history of measurement in psychology, however, experienced readers may hesitate to make this assumption.

We are indebted to S. Rains Wallace for the suggestion that a spurious factor may be present in studies which employ interviews or the extensive interpretation of questionnaires for determining morale in groups and relating it to the rated efficiency of the group.[5] It is not always clear whether the interviewers or interpreters had foreknowledge of the rated efficiency and thus might have contaminated their attitude measures.

Another procedural defect of some industrial studies is the use of self-reports or similar criterion data rather than independently obtained measures. Thus, in studies of absence and of turnover, respondents may be asked to tell how many times they have been absent in the last six months or how many jobs they have held in the last five years. Absence material, at least, could be collected in a more straightforward and probably more valid manner by looking at the individual's attendance record unless an attempt is being made to preserve anonymity.

Sometimes, furthermore, the data collection is unsupervised, or is conducted incidentally to some other operation within the company. While there is no intrinsic objection to such projects, experience indicates that they frequently suffer from errors and extraneous factors and the reader should be warned of their tentative nature.

Finally, an important procedural issue concerns the practical problem of whether or not to identify the individual respondents. If we are to relate employee attitudes to measures or estimates of individual performance, it is necessary to be able to identify individual workmen. Frequently this means that the subject is required to sign the questionnaire or is interviewed individually. Although this is a crucial problem it has not been widely investigated or, sometimes, even considered. Thus it merits extended discussion.

We have located seven studies in industrial settings which bear upon the question of the influence of anonymity upon morale or job satisfaction (**3**, **14**, **17**, **24**, **29**, **45**, **58**). Two studies from the Survey Research Center, University of Michigan, compare responses made by employees to morale items when the items are contained in an anonymous questionnaire and when they are part of an interview. In Kahn's (**29**) study, comparisons were available for 65 items which were worded identically and were presented in the same context and sequence. In general, his findings support the conclusion that identification, at least via interview, produces differential responses. For example, the questionnaire items elicited more expressions of criticism and dissatisfaction and more extreme responses. A somewhat similar study by Metzner and Mann (**45**), although much more limited, gave comparable results. Both studies contain possible defects in the matching of respondents as well as insufficient attention to interviewer differences.

---

[5] S. R. Wallace. Personal communication. October, 1954.

A related study by Wedell and Smith (58) compared interviewer ratings of responses and coded ratings of the interview protocols with questionnaire responses of approximately 200 employees of a chemical company on three attitude questions. Interviewer ratings of responses differed significantly from questionnaire responses on the two questions dealing with attitude toward company and toward job. Coded ratings of the interview protocols differed significantly from questionnaire responses on the item dealing with attitude toward the company. In each instance of significant findings the interviews produced the most favorable attitude ratings as compared to the employees' actual questionnaire responses. The most interesting finding was the differences among interviewers. By and large, the more experienced and better trained interviewers showed the greatest discrepancies! The interview versus questionnaire methodological studies do not, of course, provide a crucial test of the anonymity factor since many other variables may be operating to confound the results (**27**).

In an abstract, Evans (**14**) reports a study of salaried employees which he says "provided for testing the hypothesis that employee anonymity should be preserved in management-conducted surveys." One major finding of his early analysis indicated that "employees are not particularly concerned about anonymity." He concludes that "the unwarranted popular assumption that employees must be anonymous is open to serious question." The brevity of the report makes it difficult to evaluate the significance of this conclusion.

In an early investigation of the problem in an industrial setting, Brayfield (**3**) compared the distribution of scores on a job satisfaction questionnaire obtained from female stenographers employed in private industry with those for a similar group employed in state civil service. The members of the industrial group identified themselves under good rapport conditions and the civil service group was anonymous. It was apparent by inspection that the two groups gave similar responses. The results are inconclusive particularly since the groups were not working under strictly comparable conditions.

Gadel and Kriedt (**17**) report briefly upon a test of the effect of anonymity upon employee attitude questionnaires. They found that the distributions of answers obtained under the differing conditions of anonymity and identification were almost identical for several groups of employees. The small differences present were inconsistent from group to group and from item to item.

In a carefully controlled study, Hamel and Reif (**24**) administered a 65-item employee attitude questionnaire to two random samples of all employees in a large department store. The members of one sample remained anonymous; those in the second sample identified themselves by name and department number. The mean scores for the two groups were not significantly different. When the group responses to each individual item were compared only two item responses in 65 were significantly different. For the number of comparisons involved the authors state that "this finding is slightly smaller than expected."

The results from studies outside the industrial situation using such diverse materials as attitude scales, opinion polls and ballots, and personality tests are equivocal. Hyman

(27, p. 185), in a considered treatment of the issue, says "the foregoing discussion should serve to make clear the complexity in estimating the nature and direction of effects due to identification or anonymity of the respondent." Our own review of the dozen or so relevant nonindustrial investigations along with the industrial studies already cited leads us to agree with Kahn's (29, p. 8) conclusion that "the studies on the effect of anonymity of response conducted during the past two decades compel us to conclude that there is no predictable effect of anonymity *per se*."

We would sound the caution that, in any study of the problem, it is necessary to differentiate literal anonymity from psychological anonymity, as Hyman (27) has suggested, since even the questionnaire studies usually require identification at least to the extent of naming one's department, and sometimes even more extensive data are called for. These requirements and situational factors may influence responses to what the investigator considers to be anonymously answered materials.

We do hazard the opinion, however, that in a situation where rapport has been carefully nurtured in all steps of the investigation, the identification of *questionnaire* materials will not necessarily result in serious distortion of responses. Yet the dilemma remains. The investigator who asks for identification on the questionnaire, and perhaps especially the investigator who relies upon the interview, risks distortion in responses, while the one who does not ask for identification is unable to make certain crucial analyses. The ethical and practical consequences of coding and disguised identification of respondents are such as to render that particular solution debatable.

General Problems of Analysis and Design

We shall be concerned here with a discussion of scaling and statistical techniques commonly in use in psychology, with the general problem of operational definitions as applied to the specific area of industrial research, and with the validity of certain group measures.

Statistical and scaling techniques. In the typical industrial study the psychologist asks a group of workers to fill out a questionnaire which is designed to measure "morale," he ranks the same employees on some aspects of their job performance, and he then runs a t test or a product-moment correlation between morale, on the one hand, and performance on the other. We believe that too little recognition has been given to the assumptions about the nature of the data which are built into these measurements and into the statistical tests.

Thus, for example, the typical morale scale makes the assumption that a unidimensional continuum is being measured and that the items of the scale are equally important in contributing to the individual's position on the continuum. Except for item analysis in the initial construction of the scales, it is doubtful whether the validity of these assumptions is ever tested and item analysis does not guarantee unidimensionality, nor is it often used for differential weighting of items. The use of mean scores and of product-moment correlation adds such further assumptions as that the intervals between points on the scale are of equal magnitude, that the parent population is normally distributed on the variables concerned, that homoscedasticity exists in the case of the Pearsonian coefficient, and so on. To the extent that these assumptions are violated, the results become ambiguous. It is doubtful if the validity of the assumptions is routinely tested.

However, since these comments hold true generally for much research in all fields of psychology, it would be unreasonable to expect a different type of analysis among observers studying morale.

Measurement of morale. McNemar (44) has commented at length upon the tendency for attitude researchers to "measure" a concept by applying a particular label (say conservatism-radicalism) to a set of questions without concern for other operational definitions of the same concept. "Morale" has probably been as greatly subject to this procedure as any other concept in psychology today. The lack of consistent definition of the term "morale" has led us to title this article "Employee *attitudes* and employee performance" in the hope that the term "attitude" provides the general and ambiguous connotation that is required to describe measurement in this field.

One likely cause of this proliferation of operational definitions is that morale is a global concept. Researchers tend to consider any area of satisfaction as important for the employee's over-all morale, for his satisfaction with his job and his company. This has led to attempts, logically and through factor analysis, to specify particular aspects of morale. Frequently subareas of morale are identified, such as satisfaction with the work group, satisfaction with pay, with supervision, with promotion, with type of work, with physical aspects of the job, or with any other component of the working environment. Questions are then developed to provide an indication of the degree of employee satisfaction in each of these areas.

This move toward increased specification of the components of the concept, and toward the development of unidimensional scales probably should be applauded. However, a problem arises of how to use the subscales. Frequently the items within a subscale are summed to provide a score on the scale, and then the individuals' positions on the subscales are summed to provide an "index" of over-all morale. In such an event, the subscales constitute simply a means of weighting items, and subscales are assumed to be equally important in producing whatever overall morale is supposed to be. The possibility remains that a more fruitful method of analysis would be to consider the subscales independently or as configurations rather than to combine them additively.

The reason usually given for using multi-item scales in measuring morale is the same as that given for using many items in tests of ability: the greater reliability of longer scales. It is assumed that increasing the number of attitude items increases the likelihood that respondents will be ordered in the same way on a second administration of the questionnaire. Yet the analogy to ability tests is not necessarily a good one. While there may be an infinite population of arithmetic problems, for example, from which test items may be chosen, the inclusion of additional items in attitude scales is likely to introduce new attitude dimensions. Unless these dimensions are highly correlated, the scores on the questionnaire may be ambiguous. For example, a person could achieve a moderate morale score by indicating moderate satisfaction with his pay and with the type of work he is doing, or by indicating great satisfaction with the type of work he is doing and great dissatisfaction with his pay. It is not unlikely that these different patterns of response would be differently related to job performance, to absence, and to turnover.

Concern with unidimensionality, rather than with reliability as traditionally conceived, would lead to the use of such scaling techniques as Guttman's, Loevinger's, or Coombs's, or to the use of single items. It is interesting that many of the studies we have reported above have made at least partial use of the single-item approach (**32, 35, 46, 48, 60, 61**).

Use of group measurements. Frequently, members of a work group or a department fill out morale questionnaires individually, but have group, rather than individual, performance or productivity records. In such situations it is customary to test the significance of the difference between the morale scores of the two groups to determine whether the higher producing group also has higher morale. We shall comment upon only one aspect of this procedure by pointing out that a relationship which exists at the individual level between satisfaction and productivity may be obscured when the individuals are lumped together. There may be a positive relationship between satisfaction and performance within each group even though the two groups as a whole do not differ significantly with regard to mean satisfaction. Such a relationship obviously is not revealed in this type of analysis and lack of individual performance records makes the appropriate analysis impossible.

The degree to which these and other methodological issues may have beclouded the relationships between employee attitudes, performance, absenteeism, and employment stability is an open question. Certainly, as must be obvious to the careful reader, the studies we have reviewed are subject to some or, occasionally, most of the shortcomings described above. However, the scarcity of relationships, either positive or negative, demonstrated to date even among the best designed of the available studies leads us to question whether or not methodological changes alone would lead to a substantial increase in the magnitude of the obtained relationships. We are led, then, from consideration of the current status of research on this topic to a discussion of the relationships on the conceptual level. Much of what we will say has previously been elaborated by Katz and Kahn (**33**) and by Morse (**47**).

THEORETICAL CONSIDERATIONS

Morale as an Explanatory Concept in Industrial Psychology

One principal generalization suffices to set up an expectation that morale should be related to absenteeism and turnover, namely, that organisms tend to avoid those situations which are punishing and to seek out situations that are rewarding. To the extent that worker dissatisfaction indicates that the individual is in a punishing situation, we should expect dissatisfied workers to be absent more often and to quit the job at a higher rate than individuals who are satisfied with their work. Since the general proposition about the effects of reward has received a great amount of verification in psychology, it is not strange that it has been carried to the analysis of absenteeism and turnover.

A plausible connection between satisfaction and performance on the job is less obvious. Let us consider specifically the possible relationship between satisfaction and productivity. Under conditions of marked dissatisfaction it is likely that low productivity may serve as a form of aggression which reflects worker hostility toward management. But the hypothesis that production should increase monotonically with increases in satisfaction apparently rests on the assumption that the worker will demonstrate his gratitude by increased output, or that the increased satisfaction frees certain creative energies in the worker, or that the satisfied employee accepts management's goals, which include high production.

In any event, it is commonly hypothesized that, whatever the causes, increased satisfaction makes workers more motivated to produce. Given this condition, it should follow that increased productivity can be attained by increasing worker satisfaction. We are going to advance the proposition that the motivational structure of industrial workers is not so simple as is implied in this formula. We feel that research workers have erred by overlooking individual differences in motivations and perceptions because of their concern with discovering important and applicable generalizations. Most of what follows is an effort to point out areas in which differences between workmen may make a difference in their adjustment to the situation.

At the outset let us make it clear that we expect the relation between satisfaction and job performance to be one of concomitant variation, rather than cause and effect. It makes sense to us to assume that individuals are motivated to achieve certain environmental goals and that the achievement of these goals results in satisfaction. Productivity is seldom a goal in itself but is more commonly a means to goal attainment. Therefore, as G. M. Mahoney has suggested,[6] we might expect high satisfaction and high productivity to occur together when productivity is perceived as a path to certain important goals and when these goals are achieved. Under other conditions, satisfaction and productivity might be unrelated or even negatively related.

In the light of this consideration, we shall center our discussion on an analysis of industrial motivation as it relates specifically to employee satisfaction and to productivity.

For the sake of convenience we may distinguish between threats and rewards as incentives to productivity. Goode and Fowler (**21**) have described a factory in which morale and productivity were negatively related but productivity was kept high by the continuance of threats to workers. Here the essential workers —people with considerable skill— were marginal to the labor force because of their sex or because of physical handicaps. Since the plant was not unionized, it was possible for management to demand high productivity from these workers on threat of discharge. This meant that the workers, although most dissatisfied with their jobs, produced at a very high rate because of the difficulty they would face in finding another position should they be discharged.

There is little doubt that threat was widely used as a motivating device in our own society in the past and is presently used in more authoritarian societies. However, it is doubtful if any great amount of at least explicit threat is currently used by industries in this country in efforts to increase productivity or reduce absenteeism. First of all, considerable change has occurred in management philosophy over the past fifty years, and such tactics are repugnant to many industrial concerns. Secondly, the growth of unions has virtually outlawed such tendencies except in small, semi-marginal industries which are not unionized.

---

[6] G. M. Mahoney. Personal communication. March, 1953.

Threats of discharge, then, probably do not operate as incentives unless the worker falls considerably below the mean in quantity and/or quality of output. For a number of reasons management has settled upon rewards for motivating workers to produce, including such tangible incentives as increased pay and promotion, as well as verbal and other symbolic recognition. Let us examine whether this system of rewards actually provides motivation for increased productivity by the worker.

It is a commonplace observation that motivation is not a simple concept. It is a problem which may be attacked at a number of different levels and from many theoretical points of view. Whatever their theoretical predilection, however, psychologists generally are agreed that human motivation is seldom directed only toward goals of physical well-being. Once a certain minimum level of living has been achieved, human behavior is directed largely toward some social goal or goals. Thus, in our own society, goals such as achievement, acceptance by others, dominance over others, and so on, probably are of as great concern to the average workman as the goals of finding sufficient food and shelter to keep body and psyche together.

We assume that social motives are of considerable importance in industry. We assume, further, that the goals an individual pursues will vary, depending upon the social systems within which he is behaving from time to time. Most industrial workers probably operate in a number of social systems. Katz and Kahn (**33**) suggest four such systems: first, the system of relations outside the plant; and, within the plant, the systems of relationship with fellow workers on the job, with members of the union, and with others in the company structure. We may ask whether job performance, and particularly productivity, is a path to goal achievement within these various sets of social relations.

Outside the plant. It is often argued that any worker who is motivated to increase his status in the outside community should be motivated toward higher productivity within the plant. Productivity frequently leads directly to more money on the job, or involves movement to jobs with higher prestige or with authority over others. If productivity does result in such in-plant mobility, increased output may enable the individual to achieve a higher level of living, to increase his general status in the community, and to attempt such social mobility as he may desire. In this way productivity may serve as a path to the achievement of goals outside the plant.

The operation of this chain of relationships, however, depends not only upon the rewards given the high producer, but also upon the original motivation of the workman to increase his status position in the outside community. The amount of status motivation among production-line employees is open to question. Certainly the findings of Warner (**57**), Davis and Gardner (**12**), and others (**6**, **11**, **13**), indicate that there are systematic differences in the goals which are pursued in the different segments of our society. It is not impossible that a very large proportion of America's work force is only minimally motivated toward individual social achievement. The assumption that such a motivation does exist may reflect in considerable part a projection of certain middle-class aspirations onto working-class employees.

Furthermore, it is not unlikely

that the reference group against which an individual workman evaluates his success may be only a segment of the community, rather than the community as a whole. An individual whose accomplishments are modest at best when compared with the range of possible accomplishments in the community may have a feeling of great accomplishment when he compares his achievements with those of others in his environment. If this is true, and if he desires to continue to operate within this segment of society, any further increase in rewards within the plant might lead to his exclusion from personally important groups outside the plant rather than to increased prestige in such groups.

Finally, there are many goals outside the industrial plant which may be socially rewarding to the individual and which require only minimal financial and occupational rewards inside the plant. Active participation in veterans' organizations, in churches, in recreational programs and similar activities may be and frequently are carried out by individuals at all positions in the industrial hierarchy. As a matter of fact, to the extent that the individual receives extensive social rewards from such activities he may have only slight interest in his work on the job, and he may continue to remain in industry only to maintain some minimum economic position while carrying out his outside functions. For such an individual, high productivity may lead to *no* important goals.

Relations with other workers in the plant. The studies by Elton Mayo and his associates (**43**, **50**, **51**) introduced the work group into the analysis of industry, and a wealth of subsequent investigations have confirmed the importance of on-the-job groups. Throughout these studies has run the observation that members of the work group develop group standards of productivity and attempt to force these standards upon those workmen who deviate. Thus, in the Bank Wiring Room (**51**) it was the socially maladjusted individual, the deviant from the work group, who maintained a level of production above that of the group even though his native ability was considerably below that of many of the others.

Mathewson's (**42**) classic study of restriction of output among unorganized workers was an early demonstration of the operation of group norms.

Schachter and associates (**52**) have conducted an experiment which indicates that in cohesive groups an individual's productivity may be either raised or lowered, depending upon the kind of communications directed toward him by congenial co-workers. In an actual factory setting, Coch and French (**8**) presented existent groups with evidence that a change in job methods and in productivity was necessary if the factory was to remain in a favorable position relative to other, competing factories. These groups, through group discussion, arrived at a decision as to the proper job set up, and modified the group judgment of "fair" output markedly upward.

There is evidence, then, that the level of performance on the job frequently depends upon a group norm, and that performance level may be changed by changing the group norm in a direction desired by management. This change in the norm probably results from a conviction among the workers that higher production is in their own interest as well as management's, i.e., that their interests and management's interests coincide. This raises the perplexing question of whether, with regard to productivity, the interests of management and labor do, in fact, coincide.

Management, presumably, is interested in higher production as a way of reducing the ratio of cost to output, and thereby bettering management's financial and competitive position. In an expanding market, the argument goes, this makes possible the expansion of the company, increased wages, a larger labor force, and general prosperity not only for the corporation but for the employees as well.

The case may not be so attractive to the workers, especially when the market is not expanding and demand for the product is constant, nearly constant, or declining. In this event, higher productivity per worker means that fewer people are required for the same level of output, or that fewer hours are worked by the same number of workers. In either case, many workers may lose, rather than gain, by the increase in productivity. It may be argued that in normal times such individuals usually find fairly rapid employment in some other segment of the economy. However true this may be, from the viewpoint of the individual workman this involves a considerable disruption in working habits and in his social life in general, and is to be avoided wherever possible. Viewed in this light the interests of management and labor are inimical.

As psychologists we steer clear of such arguments. But we should be sensitive to the fact that the question is a debatable one, that a final decision will probably rest upon values rather than data, that each side is capable of convincing arguments, and that the perception of a certain inevitable conflict of interests between the two groups is honestly and intelligently held by many people. We should also recognize that any reduction in work force after a joint labor-management effort to increase productivity will likely be interpreted as resulting from the increased productivity, and may lead to a future avoidance not only of high productivity levels but also of labor-management cooperation.

At any rate, we often find that individual workers interpret higher productivity as counter to the interests of the employees. To the extent that this perception constitutes a group norm, such motives as are rewarded through the individual's social relationships with other workmen may be blocked by increased productivity. In such cases, productivity may serve as a path to certain goals, but as a block to social acceptance.

The union structure. One system of relationships of considerable importance in many industrial concerns is the union. In many companies much of what was said in the preceding section may be extended to refer also to the relations of the worker in the system of social relations within the union.

In some plants high productivity is not a deterrent to active union participation. Nevertheless, it probably is true that productivity is seldom a prerequisite for advancement within the union hierarchy. If the individual is oriented toward the union structure, it is unlikely that high productivity will serve as a path to such goals, whatever its effect on other goals he may pursue.

The company structure. We have indicated above that many of the worker's social motives outside the plant, as well as his desires for in-plant associations with fellow workmen and within the union, may be only slightly affected by increases in productivity and sometimes may be blocked by increased productivity.

The apparent range of goals that a worker may have is so wide that productivity may be a path to only a few of them.

However, workers are often motivated toward goals within the plant such as turning out a quality product, higher wages, and promotion. Let us examine the relationship between satisfaction and productivity for workers who are motivated toward these in-plant goals.

At the start it is evident that productivity and quality are sometimes mutually exclusive. If the individual must concentrate on maintaining high quality work, speed of production probably plays a secondary role. Conversely, if he must emphasize speed, quality often must be reduced to some degree. The speed-quality dilemma is sometimes resolved by making the individual work units so routine and concerned with such minute changes in the material that increased speed will not affect the quality of the product. However, if a worker is more highly motivated when he is performing some meaningful job, the above procedure may be resolving one dilemma by raising another. At any rate, the artisan, motivated toward the goal of quality, may be highly satisfied with his job while turning out a very limited number of finished pieces per unit of time. If he is forced to increase productivity and lower in some measure the quality, we might expect his satisfaction to decrease. For such a person satisfaction and productivity would be negatively related.

Consider now the individual who is motivated toward higher wages and promotion. While these rewards may not be exclusively dependent upon job performance, at the same time productivity and other aspects of performance often are weighted heavily at promotion time in most companies. In other words, productivity and other aspects of job performance constitute a path to the goal of promotion and wage increases.

Now it is likely that people with aspirations to change position in the company structure will often be quite dissatisfied with their present position in the company. Aspiration to move within a system implies not only a desire for some different position in the future, but some degree of dissatisfaction with the position one is presently occupying. The amount of dissatisfaction probably depends upon the length of time the individual has occupied this position. Thus, although productivity may be a path to the goal, failure to achieve the goal to date may result in dissatisfaction and the high producer may be less satisfied than the low producer.

Evidence sustaining this point of view is to be found in Katz and associates' (**34**) report of a large insurance company in which the best, most productive workers were also considerably more critical of company policy than were less productive workers. S. Lieberman reports a similar finding in a large appliance factory.[7] A year after all workers in the factory had filled out a questionnaire, Lieberman compared the earlier responses of those who had been promoted to foreman with a matched group of workers who were not promoted. Those promoted had been significantly less satisfied with company practices at the earlier time than had the control group.

Once again the question arises as to what is meant by satisfaction. It may be that extremely high satisfaction is indicative of a certain amount of complacency, a satisfaction with the job as it is, which may be only

[7] S. Lieberman. Personal communication. July 15, 1954.

slightly related to job performance, if it is related at all. On the other hand, individuals who are highly motivated may perceive productivity as a path to their goals, but may also be more realistically critical of whatever deficiencies exist within the organization. They may feel, in addition, that their output is not being rewarded as rapidly as it deserves.

Implications for Future Research

We have arrived at two conclusions: first, that satisfaction with one's position in a network of relationships need not imply strong motivation to outstanding performance within that system, and, second, that productivity may be only peripherally related to many of the goals toward which the industrial worker is striving. We do not mean to imply that researchers should have known all along that their results would be positive only infrequently and in particular circumstances. We have been operating on the basis of hindsight and have attempted to spell out some of the factors which may have accounted for the failure of industrial investigators to find positive relationships in their data.

However, certain implications seem logical from the foregoing sections of this report. Foremost among these implications is the conclusion that it is time to question the strategic and ethical merits of selling to industrial concerns an assumed relationship between employee attitudes and employee performance. In the absence of more convincing evidence than is now at hand with regard to the beneficial effects on job performance of high morale, we are led to the conclusion that we might better forego publicizing these alleged effects.

The emphasis on predicting job performance, and particularly productivity, rests upon the acceptance of certain values. That is, the many studies that have used productivity as the criterion to be predicted have been performed because productivity has direct economic value to industry, and, presumably, to society at large. But the fact that it has economic value does not mean that job performance is the only, or even the most important, aspect of organizational behavior. From the viewpoint of studying, analyzing, and understanding the industrial setting and individual reactions thereto, productivity and other aspects of job performance may be only one of several important factors. It would seem worthwhile to study the causes, correlates, and consequence of satisfaction, per se. It seems possible, for example, that conditions conducive to job satisfaction will have an effect on the quality of the workman drawn into the industry, the quality of job performance, and the harmony of labor-management relations. Such potential correlates, among others, merit exploration.

Another potentially fruitful approach involves studying the differential effect of particular kinds of management practices upon the attitudes and performances of workers with different motives, aspirations, and expectations. The appropriate questions may concern how, for particular workers, productivity comes to be perceived as instrumental to the achievement of some goals but not others, while for other workers a different perception develops.

The experimental approach has largely been neglected in this area of industrial research, yet the control of variables that it provides seems essential to the development and refinement of our knowledge in the area. Certainly, where experimentation has been used, as by Schachter and associates (52) and by Coch and

French (8), the results have been both enlightening for the understanding of present problems and encouraging for its future application. As our concepts become increasingly precise, we may expect an increased use of experimentation both within the industrial setting and in the laboratory.

Perhaps the most significant conclusion to be drawn from this survey of the literature is that the industrial situation is a complex one. We have suggested that an analysis of the situation involves analysis not only of the individual's relation to the social system of the factory, the work group, and the union, but the community at large as well. It is important to know what motives exist among industrial workers, how they are reflected in the behavior of the workers, and how the motives develop and are modified within the framework of patterned social relationships in the plant and in the larger community.

We seem to have arrived at the position where the social scientist in the industrial setting must concern himself with a full-scale analysis of that situation. Pursuit of this goal should provide us with considerable intrinsic job satisfaction.

REFERENCES

1. BAXTER, B., TAAFFE, A. A., & HUGHES, J. F. A training evaluation study. *Personnel Psychol.*, 1953, **6**, 403–417.
2. BERNBERG, R. E. Socio-psychological factors in industrial morale: I. The prediction of specific indicators. *J. soc. Psychol.*, 1952, **36**, 73–82.
3. BRAYFIELD, A. H. *The interrelation of measures of ability, aptitude, interests, and job satisfaction among clerical employees.* Unpublished doctor's dissertation, Univer. of Minnesota, 1946.
4. BRAYFIELD, A. H., & ROTHE, H. F. An index of job satisfaction. *J. appl. Psychol.*, 1951, **35**, 307–311.
5. BRODY, MILDRED. *The relationship between efficiency and job satisfaction.* Unpublished master's thesis, New York Univer., 1945.
6. CENTERS, R. *The psychology of social classes.* Princeton: Princeton Univer. Press, 1949.
7. CHASE, F. S. Factors for satisfaction in teaching. *Phi Delta Kappan*, 1951, **33**, 127–132.
8. COCH, L., & FRENCH, J. R., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.
9. COMREY, A. L., PFIFFNER, J. M., & BEEM, HELEN P. Factors influencing organizational effectiveness. I. The U. S. Forest Survey. *Personnel Psychol.*, 1952, **5**, 307–328.
10. COMREY, A. L., PFIFFNER, J. M., & BEEM, HELEN P. Factors influencing organizational effectiveness. II. The Department of Employment Survey. *Personnel Psychol.*, 1953, **6**, 65–79.
11. DAVIS, A. *Social class influences upon learning.* Cambridge: Harvard Univer. Press, 1948.
12. DAVIS, A., GARDNER, B. B., & GARDNER, MARY R. *Deep south: A social and anthropological study of caste and class.* Chicago: Univer. of Chicago Press, 1941.
13. ERICSON, MARTHA C. Social status and child rearing practices. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947. Pp. 494–501.
14. EVANS, C. E. Item structure variation as a methodological problem in an employee survey. *Amer. Psychologist*, 1949, **4**, 280. (Abstract)
15. FESTINGER, L., & KATZ, D. (Eds.) *Research methods in the behavioral sciences.* New York: Dryden Press, 1953.
16. FRIESEN, E. P. The incomplete sentences technique as a measure of employee attitudes. *Personnel Psychol.*, 1952, **5**, 329–345.
17. GADEL, MARGUERITE S., & KRIEDT, P. H. Relationships of aptitude, interest, performance, and job satisfaction of IBM operators. *Personnel Psychol.*, 1952, **5**, 207–212.
18. GARRETT, H. E. *Statistics in psychology and education.* New York: Longmans, Green, 1947.
19. GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology.* New York: McGraw-Hill, 1948.

20. Giese, W. J., & Ruter, H. W. An objective analysis of morale. *J. appl. Psychol.*, 1949, **33**, 421–427.

21. Goode, W. J., & Fowler, I. Incentive factors in a low morale plant. *Amer. sociol. Rev.*, **1949**, **14**, 618–624.

22. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

23. Habbe, S. Job attitudes of life insurance agents. *J. appl. Psychol.*, 1947, **31**, 111–128.

24. Hamel, L., & Reif, H. G. Should attitude questionnaires be signed? *Personnel Psychol.*, 1952, **5**, 87–91.

25. Hill, J. M. M., & Trist, E. L. A consideration of industrial accidents as a means of withdrawal from the work situation. *Hum. Relat.*, 1953, **6**, 357–380.

26. Houser, J. D. *What the employer thinks.* Cambridge: Harvard Univer. Press, 1927.

27. Hyman, H. H. *Interviewing in social research.* Chicago: Univer. of Chicago Press, 1954.

28. Jahoda, Marie, Deutsch, M., & Cook, S. W. (Eds.) *Research methods in social relations.* New York: Dryden Press, 1951.

29. Kahn, R. L. *A comparison of two methods of collecting data for social research: The fixed-alternative questionnaire and the open-ended interview.* Unpublished doctor's dissertation, Univer. of Michigan, 1952.

30. Katz, D. Morale and motivation in industry. In W. Dennis (Ed.), *Current trends in industrial psychology.* Pittsburgh: Univer. of Pittsburgh Press, 1949. Pp. 145–171.

31. Katz, D., & Hyman, H. Industrial morale and public opinion methods. *Int. J. Opin. Attit. Res.*, 1947, **1**, 13–30.

32. Katz, D., & Hyman, H. Morale in war industry. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947. Pp. 437–447.

33. Katz, D., & Kahn, R. L. Some recent findings in human relations research in industry. In G. E. Swanson, T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1952. Pp. 650–665.

34. Katz, D., Maccoby, N., & Morse, Nancy. *Productivity, supervision and morale in an office situation.* Univer. of Michigan: Survey Research Center, 1950.

35. Katz, D., Maccoby, N., Gurin, G., & Floor, L. G. *Productivity, supervision and morale among railroad workers.* Univer. of Michigan: Survey Research Center, 1951.

36. Kendall, Patricia, L., & Lazarsfeld, P. F. Problems in survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research.* Glencoe, Ill.: Free Press, 1950.

37. Kerr, W. A. On the validity and reliability of the job satisfaction Tear Ballot. *J. appl. Psychol.*, 1948, **32**, 275–281.

38. Kerr, W. A. Summary of validity studies of the Tear Ballot. *Personnel Psychol.*, 1952, **5**, 105–113.

39. Kornhauser, A., & Sharp, A. Employee attitudes: suggestions from a study in a factory. *Personnel J.*, 1932, **10**, 393–401.

40. Kristy, N. F. *Criteria of occupational success among post office counter clerks.* Unpublished doctor's thesis, Univer. of London, 1952.

41. Lawshe, C. H., & Nagle, B. F. Productivity and attitude toward supervisor. *J. appl. Psychol.*, 1953, **37**, 159–162.

42. Mathewson, S. B. *Restriction of output among unorganized workers.* New York: Viking Press, 1931.

43. Mayo, E. *The social problems of an industrial civilization.* Cambridge: Graduate School of Business Administration, Harvard Univer., 1945.

44. McNemar, Q. Opinion-attitude methodology. *Psychol. Bull.*, **1946**, **43**, 289–374.

45. Metzner, Helen, & Mann, F. A limited comparison of two methods of data collection: the fixed alternative questionnaire and the open-ended interview. *Amer. sociol. Rev.*, 1952, **17**, 486–491.

46. Metzner, Helen, & Mann, F. Employee attitudes and absences. *Personnel Psychol.*, 1953, **6**, 467–485.

47. Morse, Nancy C. *Satisfactions in the white-collar job.* Univer. of Michigan, Survey Research Center, 1953.

48. Mossin, A. C. *Selling performance and contentment in relation to school background.* New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1949.

49. Raube, S. A. *Experience with employee attitude surveys* (Studies In Personnel Policy, No. 115). New York: National Industrial Conference Board, 1951.

50. Roethlisberger, F. J. *Management and morale,* Cambridge: Harvard Univer. Press, 1943.

51. Roethlisberger, F. J., & Dickson, W. J. *Management and the worker.* Cam-

bridge: Harvard Univer. Press, 1939.
52. Schachter, S., Ellertson, N., McBride, D., & Gregory, D. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, **4**, 229–238.
53. Stagner, R., Flebbe, D. R., & Wood, E. F. Working on the railroad: a study of job satisfaction. *Personnel Psychol.*, 1952, **5**, 293–306.
54. Thorndike, R. L. *Personnel selection.* New York: Wiley, 1949.
55. Van Zelst, R. H., & Kerr, W. A. Workers' attitudes toward merit rating. *Personnel Psychol.*, 1953, **6**, 159–172.
56. Viteles, M. S. *Motivation and morale in industry.* New York: Norton, 1953.
57. Warner, W. L., & Lunt, P. S. *The social life of a modern community.* New Haven: Yale Univer. Press, 1941.
58. Wedell, C., & Smith, K. U. Consistency of interview methods in appraisal of attitudes. *J. appl. Psychol.*, 1951, **35**, 392–396.
59. Weitz, J., & Nuckols, R. C. The validity of direct and indirect questions in measuring job satisfaction. *Personnel Psychol.*, 1953, **5**, 487–494.
60. Weschler, I. R., Kahane, M., & Tannenbaum, R. Job satisfaction, productivity and morale: a case study. *Occupational Psychol.*, 1952, **26**, 1–14.
61. Wickert, F. R. Turnover, and employees' feelings of ego-involvement in the day-to-day operations of a company. *Personnel Psychol.*, 1951, **4**, 185–197.
62. Yoder, D., Heneman, H., Jr., & Cheit, E. F. *Triple audit of industrial relations.* Minneapolis: Univer. of Minnesota Press, 1951.

Received July 27, 1954.

Psychological Bulletin
Vol. 52, No. 5, 1955

THE "POST-MORTEM" TESTING OF EXPERIMENTAL COMPARISONS

RICHARD B. McHUGH
Iowa State College

AND

DOUGLAS S. ELLIS
Air Force Personnel and Training Research Center, Denver

The advantage of using an efficient experimental design which permits an over-all analysis of variance F test has been frequently pointed out, for example, in the recent texts of Johnson (3), Lindquist (5), and Walker and Lev (8). As illustration, a portion of the reading experiment data of Burt and Lewis (1) are presented[1] in Table 1. Evidently the four *remedial instruction methods*—Alphabetic, Kinesthetic, Phonic, and Visual—have significantly different effects at the 1% level.

TABLE 1

Analysis of Variance of Reading Improvement Scores from Burt and Lewis Data (1)

Source	*df*	SS	MS	*F*	$F_{1\%}$
Methods of Remedial Instruction	3	880.1	293.4	9.31	4.46
Methods of Original Teaching	3	230.6	76.9	2.44	4.46
Interaction	9	764.9	85.0	2.70	3.01
Error	32	1007.6	31.5		

TABLE 2

Totals and Means for the Four Methods of Remedial Instruction

Methods of Remedial Instruction	Total	Mean	Number of Cases
Alphabetic	$T_1=1267.3$	$M_1=105.61$	$N_1=12$
Kinesthetic	$T_2=1347.2$	$M_2=112.27$	$N_2=12$
Phonic	$T_3=1247.2$	$M_3=103.93$	$N_3=12$
Visual	$T_4=1368.6$	$M_4=114.05$	$N_4=12$

However, typically the investigator may not wish to stop with the analysis of Table 1. He may wish to go on to examine Table 2. That is, for each significant over-all F emerging, he frequently wishes to pursue the corresponding effects further by contrasting means or totals from certain groups of treatments. For example (operating at the 1% level), since F for *remedial instruction methods* is significant,[2] then the psychologist might wish to know if the significance could be attributed in part to the difference between the Alphabetic and Phonic, T_1-T_3; or possibly to the contrast between the Kinesthetic and Visual, T_2-T_4. In fact, going beyond a simple difference of two treatments, the investigator might even wish to compare the Alphabetic and Phonic methods, *jointly*, with the Kinesthetic and Visual, *jointly*, i.e., $T_1+T_3-T_2-T_4$.

---

[1] The data are also summarized as an example in Walker and Lev (8), Chap. 14.

[2] If the interaction between *remedial instruction methods* and *original teaching methods* were significant, then an investigation of the $4\times4=16$ treatment combination would be in order; i.e., the investigator would want to examine the significance of the *remedial* effects for each level of *original teaching* effects, and vice versa.

Two possible cases arise in the selection of such experimental comparisons for statistical testing: first, the a priori situation, i.e., where the comparisons have been planned in advance of the experiment; second, the a posteriori (or "post-mortem") situation, i.e., where the comparisons have not been pre-planned, but are suggested by an examination of the data. Only for the first case are the tabular F and t values appropriate, since these values are based on statistical theory which assumes *randomly* selected observations. Therefore, the conventional t or F test may be misleading in the second case (which is the more typical and acutely frustrating case to the investigator), where the comparisons are not determined solely by the nature of the experiment and design but are dependent upon the data.

Special methods are needed in this second case because the confidence levels are upset when the results of the investigation are used to decide on the comparison to be made. The most notorious example of making tests of significance on hypotheses suggested by the data is that of waiting until the experiment is completed and then selecting the highest and lowest means for analysis by an ordinary t test (or the equivalent F test for 1 degree of freedom in the numerator). The blunder in this procedure lies in the fact that the sampling distribution of such mean differences (between largest and smallest means) is more variable than the conventional distribution. And, as indicated by Cochran and Cox (2), in general the error variances for such "post-mortem" comparisons are often considerably larger than those error variances appropriate to comparisons which have been hypothesized for test independently of the current data. The direction of the bias is clear. By using conventional (case one) test procedures in this unconventional (case two) situation, too small an error variance is used and the investigator will tend to obtain a large t or F value, yielding a spuriously significant comparison.

Various routes are open to an investigator desirous of fully examining his data yet aware of the danger of an erroneous inference. The following procedure has both the virtues of simplicity and conservatism to recommend it. Test the comparison by the ordinary numerator-one-degree-of-freedom F test (or equivalent t test). (*a*) If the comparison is not significant at the stipulated confidence level by this test, then the investigator can be sure of nonsignificance. (*b*) If the comparison is apparently significant by this test, the significance is possibly spurious. The technique illustrated below for dealing further with this (*b*) situation is due to Scheffé (7). Essentially, one simply uses the same computed F ratio but alters the apparent size of the critical region, conservatively.

For example, consider the "post-mortem" testing of the null hypothesis that the Alphabetic and Phonic effects are identical. For this one degree of freedom comparison, $C: T_1-T_3$, the sum of squares[3] is

$$SS_C=\frac{(T_1-T_3)^2}{N_1+N_3}=\frac{(1267.3-1247.2)^2}{12+12}$$
$$=16.83.$$

The mean square for this comparison

---

[3] Walker and Lev (8, p. 356), give computing instructions for finding the SS due to any comparison.

is therefore $MS_C = SS_C/1 = 16.83$. Finally, the observed F ratio is

$$F = MS_C/MS_{Error} = 16.83/31.5 = .53$$

where the denominator mean square, with $n_2 = 32df$, is taken from the analysis of variance in Table 1. The tabular 1% F value for $n_1 = 1$ and $n_2 = 32df$ is $F_{1\%,1,32} = 7.50$. Hence this comparison does not reach the 1% significance level of the conventional test. A fortiori, it is not significant, i.e., it would surely fail to reach significance if the appropriate 1% level were available. This illustrates situation *a* above.

It should be noted that instead of performing the test of significance by means of the F ratio, a general alternative is to use the corresponding t statistic. They are entirely equivalent since $F_{1,n_2} = t_{n_2}^2$, i.e., F with $n_1 = 1$ df for the numerator and n_2 df for the denominator is equal to the square of t with n_2 df. As illustration, the present comparison $C: T_1 - T_3$ can be equivalently expressed as a difference of means, $C: M_1 - M_3$, instead of as a difference of totals. Then $t = (M_1 - M_3)/SE_{M_1 - M_3}$ where the standard error[4] is given by

$$SE_{M_1-M_3} = \sqrt{MS_{Error}(1/N_1 + 1/N_3)}$$
$$= \sqrt{31.5(1/12 + 1/12)} = 2.29.$$

Hence

$$t = \frac{105.61 - 103.93}{2.29} = .73,$$

the square of which (.5329) is approximately $F = .53$. Whether t or F is used in such applications is a matter of taste, except in the case of a one-sided hypothesis where, of course, the t statistic is employed since the sign of the comparison is fundamental. Jones (4) discusses the rationale of unilateral comparisons.

As a second example of a "post-mortem" statistical test, consider the hypothesis that the effects of the Alphabetic and Phonic methods, *jointly*, is identical with the *joint* effect of the Kinesthetic and Visual treatments. For this one degree of freedom comparison, $C: T_1 + T_3 - T_2 - T_4$, the sum of squares is

$$SS_C = \frac{(T_1 + T_3 - T_2 - T_4)^2}{N_1 + N_2 + N_3 + N_4}$$
$$= \frac{(1267.3 + 1247.2 - 1347.2 - 1368.6)^2}{12 + 12 + 12 + 12}$$
$$= 844.20.$$

The mean square for this contrast is therefore $MS_C = 844.20$ and the observed F ratio is $F = 844.20/31.5 = 26.8$. Since 26.8 exceeds the 1% point of 7.50, then this comparison is apparently significant. However, this example illustrates situation *b* above; i.e., the apparent significance of this contrast may be spurious. The Scheffé modification of the significance region is therefore in order. In general, at the $\alpha\%$ significance level, it consists of using as significance point,[5] the value $(k-1)F_{\alpha,k-1,n_2}$ instead of $F_{\alpha,1,n_2}$. In this expression, k denotes the total number of treatments in the group of treatments from which the comparison has been constructed. Thus here the value $(4-1)\ F_{1\%,4-1,32}$, or $3\ F_{1\%,3,32}$ which is $(3)\ (4.46) = 13.38$, would be used instead of $F_{1\%,1,32} = 7.50$. Since the observed F of 26.8 exceeds even this modified 1% significance point of 13.38, the reality of the difference between the effects of Alphabetic-Phonic and Kinesthetic-Visual seems assured. Evidently, the significance region is altered conservatively by the Scheffé method, i.e., a larger observed F is needed to obtain significance at a stated level when the comparison is devised a posteriori (and so capitalizes on possibly fortuitously large effects) instead of a priori. This is in accord with the considerations mentioned earlier.

---

[4] Lindquist (5, p. 14), gives computing instructions for evaluating the standard error for any comparison.

[5] When the comparison of interest is a simple difference between two means, a more sensitive test due to Tukey may be employed. Tukey's test requires tables of the Studentized range instead of the F table only. An example is to be found in Scheffé (6).

REFERENCES

1. Burt, C., & Lewis, R. B. Teaching backward readers. *Brit. J. educ. Psychol.*, 1946, **16**, 116–132.
2. Cochran, W. G., & Cox, G. *Experimental designs.* New York: Wiley, 1950.
3. Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
4. Jones, L. V. Tests of hypotheses: one-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43–46.
5. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
6. Scheffé, H. An analysis of variance for paired comparisons. *J. Amer. Statist. Ass.*, 1952, **47**, 381–400.
7. Scheffé, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.
8. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

Received July 6, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 5, 1955

THE INFANTILE DISORDERS OF HOSPITALISM AND ANACLITIC DEPRESSION[1]

SAMUEL R. PINNEAU[2]
University of California

Within the last ten years, psychogenic disturbances in infancy have received increasing recognition (**34**). In this connection Spitz (**43, 44, 45**), Ribble (**38, 39**), and Fischer (**16**) have especially emphasized the importance of the mother-child relationship, and their reports have been increasingly accepted and quoted by many. At least two major textbooks in general psychology have given considerable space to these authors' works (**25, 41**). Their works receive even greater attention and acceptance in Maslow and Mittelmann's recent revision of their text for abnormal psychology (**34**), where they constitute approximately half of the chapter on disorders of infancy. A favorable reception is also apparent in articles by Geleerd (**18**), by Hartmann, Kris, and Loewenstein (**21**), and by Bowlby (**9**). Kris' evaluation of these writers' works appears to be representative of this school: " . . . it seems appropriate to state how much we owe to Margaret Ribble's own investigation . . . and to the long set of investigations on the consequences of the early institutionalization of the child, to which R. Spitz has contributed so decisively . . . " (**30**, p. 31). Less favorable, or even skeptical, appraisals may be found in reports by Jones (**29**), by Dennis (**14**), and by Orlansky (**35**). The increasing recognition given these studies and the somewhat contradictory evaluations of them indicate the need for a more complete and critical consideration of them.

As the writer has evaluated the writings of Ribble elsewhere (**36**), they will not be considered in the present review. Primary consideration will be given to the investigations which Spitz reports (**43, 44, 45, 46, 47**) as these are the most extensive and as they have been reported in the greatest detail. A brief summary of his investigations will be presented before we attempt to evaluate them. An evaluation of Fischer's work may be found in a separate section at the end of this article.

THE SPITZ PAPERS

The Spitz papers which will be considered in this article are those which have been published in the first six volumes of *The Psychoanalytic Study of the Child*. The first of these, entitled "Hospitalism. An Inquiry Into the Genesis of Psychiatric Conditions in Early Childhood," was published in 1945 in Volume I of this series (**43**). In this paper he contrasts the development of infants in two institutions which he names "Nursery" and "Foundling Home." These will be described in some detail in the next section. The report deals specifically with an analysis of the data in terms of the children's development as measured by the Hetzer-Wolf tests, a comparison of the background of the two groups, a comparison of their physical development, and a comparison of their care in the two institutions in terms of housing conditions, food, clothing, medical care, personnel, restrictions on activity, and amount of contact with their true or substitute mothers.

---

[1] The writer wishes to acknowledge the criticisms and suggestions of Harold E. Jones, which have been invaluable in the writing of this article.

[2] Part of the work on this review was done while the writer was on the staff of the University of Oregon.

Spitz's second and third articles were published in 1946 in the second volume of this series. The second article, "Hospitalism. A Follow-up Report on Investigation Described in Volume I, 1945," deals chiefly with a description of the development of the Foundling Home infants in the two years subsequent to the original study (**44**). These descriptions were based on data collected by an investigator who visited Foundling Home at four-month intervals during this time. Only 21 of the original group of 91 infants were still available for study at the end of this time. He contrasts the development of these children with the development of 122 infants from the other institution, Nursery. As only 69 infants were present in Nursery in the original study (**43**), 53 new cases had been added to the sample for this study.

His third article is entitled, "Anaclitic Depression. An Inquiry into the Genesis of Psychiatric Conditions in Early Childhood, II" (**45**). The term *anaclitic depression* refers to a psychiatric syndrome of depression which Spitz observed in some of the infants studied in the institution called Nursery. This syndrome was evident in severe form in 19 of the infants and in mild form in 26 of them. All cases manifesting the disorder had been separated from their mothers, but not all who underwent separation developed the syndrome. (The number of children who were separated from their mothers was not given.) He contrasts the reaction of the 45 Nursery children manifesting anaclitic depression with the development of the children in Foundling Home who were separated from their mothers prevalently in the sixth month. In addition he includes a review of the literature on the concept of early depression.

His fourth article, "Autoerotism. Some Empirical Findings and Hypotheses on Three of its Manifestations in the First Year of Life," was published in 1949 in Volume III–IV of *The Psychoanalytic Study of the Child* (**46**). The three manifestations of autoerotism investigated were "rocking," genital play, and fecal games. He relates these to each other and to the intensity of the mother-child relationship. He compared the incidence of these activities in three groups of subjects, those reared in Nursery, Foundling Home, and family homes. The main body of the investigation was based on 170 of the 190 infants now included in the Nursery sample.

The fifth and final article with which we will be concerned was called "The Psychogenic Diseases in Infancy: An Attempt at their Etiologic Classification" (**47**). It was published in 1951 in the sixth volume of this series. This paper presents Spitz's attempt to set up a classification of psychogenic diseases of infancy. This report adds little to our knowledge of the growth and development of the children in Nursery and Foundling Home.

NURSERY AND FOUNDLING HOME

Before turning to some of Spitz's broader conclusions, it would seem desirable to consider in a general way the circumstances which led to his subjects being reared in these institutions, the background of these children's parents, the physical characteristics of the institutions and the care received by the children while there.

Nursery is a penal institution for delinquent girls. In that institution, the girls who gave birth to a child while there, cared for their infants until they were approximately a year old. While the Nursery infants are generally described as being cared for by their mothers throughout the first year, it appears in a later report by Spitz (**45**) that a considerable proportion of them were separated from their mothers for a period of three months beginning at 6–8 months of age, for reasons not indicated. When a child had to be separated from his mother, a pregnant girl or another mother cared for him. The mothers were supervised by a head nurse and three assistants. Despite the fact that he describes most of the mothers as socially maladjusted, feebleminded, psychically defective, psychopathic, or criminal, Spitz reports that in Nursery the mother "gives the child everything a good mother does and, beyond that, everything else she has" (**43**, p. 65). The children were kept in glass-enclosed cubicles until they were six months of age, at which time they were transferred to rooms in which there were a number of children. Toys were almost always present and there was a feeling of warmth and friendliness about the institution because the mothers spent a great deal of time carrying their children, tending them, feeding them, playing with them, and chatting with each other with their babies in their arms.

In contrast to Nursery is Foundling Home. The children were taken to that institutiton because their mothers were unable to care for them outside of the institution. It appears that inability to support the children was the main reason that the mothers were unable to care for them and not because they were socially maladjusted or abnormal in any way. The infants were breast-fed by their mothers who were present in the institution for several months but, for reasons not given by Spitz, the mothers seem to have had little to do with the care of their children. The children in Foundling Home were cared for primarily by five to eight nurses who were described as unusually motherly, baby-loving women. Each of theses nurses cared for eight or more children and, apparently as a result, the children lacked human contact most of the day. These children also lived in cubicles, but in this case they were enclosed on only three sides. Bed sheets were hung over the sides of the crib so that only rarely were the infants able to see what was going on in the ward about them. Foundling Home was poorly provided for financially as compared with Nursery and, when Spitz first went to the institution, about the only objects that the children had to play with were their own hands and feet.

SPITZ'S CONCLUSIONS

From his study of these subjects, Spitz concludes that those infants who are separated from their mothers for over six weeks tend to develop psychogenic disorders. We will use the term *hospitalism* to cover the syndrome of symptoms which characterized the infants of Foundling Home, in keeping with his discussion of this syndrome in his initial article (**43**). As previously stated, the syndrome manifested by the children in Nursery who were separated from their mothers will be termed anaclitic depression. Spitz makes it evident in an article (**47**) published some five years after his initial reports (**43, 44, 45**) that hospitalism is merely an exaggerated form of the disorder of anaclitic depression; however, his original descriptions of the symptoms differ somewhat, and for the purposes of this review they will be treated separately.

Two major symptoms characterized the infants of Foundling Home. One of these, and the most prominent, was a drop in the developmental quotient (DQ) as measured by the Hetzer-Wolf baby tests. The second most prominent was a change in the infants' reactions to strangers during the last third of the first year. In contrast to their usual reaction to strangers, the children's behavior now varied from blood-curdling screams to an extreme friendliness to strangers combined with an anxious avoidance of inanimate objects (**43**). Other symptoms of the hospitalism syndrome include retardation in skeletal development, in ability to sit and walk, and in ability to develop social skills (**44**). These symptoms became increasingly marked with the length of the period of separation from the mother. He concludes that these symptoms are the result of the child being separated from his mother. While he notes that the children were deprived in other ways also, he states, "The presence of a mother or her substitute is sufficient to compensate for all the other deprivations" (**43**, p. 68).

Spitz termed the syndrome "anaclitic depression," because he considered that the clinical picture was similar to that found in depression in adults. Among the symptoms shown by the infants were the following: There was a drop in the DQ as measured by the Hetzer-Wolf baby tests. They developed a weepy, apprehensive, sad behavior in contrast to their previously happy and outgoing behavior. They lay or sat with wide-open, expressionless eyes and frozen immobile faces. They looked as if they were in a daze and apparently were not perceiving what went on in their environment. This behavior was accompanied in some by autoerotic activities in the oral, anal, and genital areas. Not every symptom was apparent in every case and it seemed that, at least in some, first one symptom dominated and then another. He comes to the same conclusion in this study that he came to in his study of Foundling Home: The cause of the symptoms is the separation of the child from his mother.

We will turn now to an evaluation of his studies to determine if his conclusion is warranted by the data he presents.

DIFFICULTIES INVOLVED IN AN EVALUATION OF SPITZ'S STUDIES

A number of features of Spitz's reports on his study of the Nursery and Foundling Home infants make an evaluation of his findings difficult. Spitz gives practically no information as to the time at which the studies took place or as to the location of the institutions; however, some information can be gleaned from his various reports (**43, 44, 45, 46, 47**). In a footnote to the first report, published in 1945, he notes that his approach to the problem of hospitalism was "... mapped out and begun in 1936 ..." (**43**, p. 56). In a report (**44**) written June 12, 1946, he noted that the institution termed Foundling Home was first visited two years previously, and that the study of the other institution, Nursery, "... now covers a period of three-and-a-half years ..." (**44**, p. 116). Judging from this information, it appears that the observation of the children began in late 1942 or early in 1943. While stating that the institutions are in two different countries in the Western Hemisphere, he does not designate the countries more specifically and, according to the magazine *Time* (**50**), has refused to do so. From his references it was possible to determine that Nursery is an institution for delinquent girls located somewhere in New York state (46, 49); however, one cannot from his writings locate Foundling Home more specifically than south of the Rio Grande. The secrecy preserved concerning social and geographic areas served by these institutions is one of the factors which make it extremely difficult to evaluate the Spitz results.

Some idea regarding the magnitude of the investigation can be gained from his third report. He says regarding children in Nursery,

> These 123 infants stayed in the nursery from their fourteenth day to the end of their first year and in a few cases up to their eighteenth month. No selection was made in the infants observed. We invariably tested and followed each child admitted to the nursery up to the day when it left. The observations took place at weekly intervals, and totalled approximately 400 hours for each child (45, p. 317).

This statement would seem to imply individual observation on each child. Spitz does not reveal the size of the population in Nursery at a specified time, except to say that when the study began there were 69 infants. On the basis of an average stay of 14 months, this would seem to imply a requirement of 460 hours of observation per week, and a staff of 12 persons. In a study of this magnitude, we would usually expect to have the staff named, and some information concerning their training and qualifications, but this is another point on which Spitz remains silent.

The Samples

One of the most important considerations in comparing two groups of subjects is an analysis of the respects in which they are different and comparable. As Spitz's conclusions are based in part on the differences between the two groups at certain ages, it would seem appropriate to consider what differences would be expected on the basis of what we know about the backgrounds of the subjects. It is important that individual cases were not excluded in the study, that it was planned so as " . . . to embrace the total population of both [institutions] (130 infants)" (43, p. 56). That selection was absent in all the reported investigations is apparent for he says in his fourth report, "As in all our investigations, the unselected total sample of the children present in the institution was observed by us and used for our study" (46, p. 88). It is regrettable that he did not note the absence or frequency of congenital abnormalities in the two institutions, the birth condition of the infants, incidence of syphilis in the mothers, or other health conditions which might be relevant, and that he did not note what if any effect such factors might have had on his findings. However, these groups can better be compared after they have been discussed separately.

The Nursery Sample

As noted earlier, the institution referred to as Nursery is a penal institution for delinquent girls (43). Some of the women were pregnant on admission. After delivery and the lying-in period, the mothers assumed the care of the infants who remained in the Nursery until the end of their first year. Spitz describes the mothers as follows:

> The mothers in Nursery came there because of a failure in social adaptation. In a large percentage of the cases this maladaptation is not severe, consisting mainly in sexual indiscretion at the wrong age. . . . We suspect that the difference between the mothers in this institution and other mothers of an urban background is one based on cultural attitudes of their immediate environment and on the diversities of their economic status (46, p. 98).

Apparently Spitz continued to add cases to his sample. In his first article

(**43**) Spitz reports that the total number in the Nursery sample was 69, in the second article (**44**) 122, in a third (**45**) 123, and in a fourth (**46**) 196. There is no indication as to the ages at which the new cases were added. It is not until the third article that the group is described as both white and colored (77 and 46 cases, respectively). Since such an obvious sampling detail was omitted, one may wonder if there were other sampling details, relevant to interpreting his findings, which were also omitted.

The Foundling Home Sample

It has been indicated that children are placed in Foundling Home because their mothers are unable to provide for them. Regarding their background Spitz says:

> A certain number of the children housed have a background not much better than that of the Nursery children; but a sufficiently relevant number come from socially well-adjusted, normal mothers whose only handicap is inability to support themselves and their children (which is no sign of maladjustment in women of Latin background) (**43**, p. 60).

While this limited amount of information makes an adequate evaluation of the children's background impossible, the mothers' inability to support their children does raise several questions. Thus, is their inability to support their children the result of being of the lowest socioeconomic group, of having limited mental abilities, or of both? Why did the responsibility of supporting the children fall to the mothers rather than to the fathers? What were the handicaps or incapacities of the fathers that made this necessary? Is it reasonable to expect that one would find in *any* foundling home that " . . . the children are of an unselected urban . . . background" (**43**, p. 58), as Spitz maintains for this group? Certainly his statement that the mothers were unable to support the infants would seem to indicate that the sample was biased in its socioeconomic level.

As in the case of the Nursery group, the actual size of the Foundling Home sample is not easy to determine. Spitz (**43**) states that the study embraced the total population of children in Foundling Home, and in the subsequent table (**43**, p. 57) gives the number of children present as 61. Later in the same article he states that there were 88 children in Foundling Home up to 2½ years of age (**43**, p. 59). In a subsequent article he refers to 91 children under 3 years of age (**44**, p. 114).

Bowlby states that Spitz, in his first article, gives "explicit data" showing that the mother-separated and control groups " . . . are of a similar social class and, as nearly as possible, spring from similar stock" (**9**, p. 20). However Spitz, in comparing the background and heredity of the two groups, contends that his data show a *marked advantage* for the Foundling Home children (**43**, p. 61, 66). The author can find no support for either of these contentions in Spitz's reports. In fact from the limited information available, it appears that the Nursery group may have the advantage.

According to Spitz the difference between the Nursery mothers and other mothers of an urban background is to be accounted for in terms of differences in their cultural backgrounds and in the diversities of their economic status (**46**, p. 98). If true, there would appear to be no basis on which to maintain that the Foundling Home children have a superior heredity since previous considerations indicate that they come from the lowest socioeconomic group.

The behavioral differences between the two groups could represent, in various combinations, differences in racial extraction (**45**) and racial mixture (**43**, **45**), negative hereditary selection (**43**), and congenital defects, as well as differences in cultural factors and econonic status (**46**). It would appear that only very tentative conclusions as to the importance of the mother-child relationship should be based on differences in behavior of two groups in which there are so many other possible factors which may affect developmental data.

The Nature of the Study

It has been noted that the conclusions which Spitz draws in his reports are based, in part, on a comparison of the Nursery with the Foundling Home infants. Spitz also draws conclusions from comparisons of the age-trends of a given group with other groups and with developmental norms.

It seems to the writer that anyone who reads Spitz's articles is likely to infer from his statements that his reports are based on longitudinal data. However, there is question as to the extent to which this is true.

First let us determine if Spitz implies that a longitudinal method was used. His description of the Foundling Home children's development follows: "Their Developmental Quotient on admission is below that of our best category [a group of children from 'professional homes'] but much higher than that of the other two [the Nursery infants and a group of infants from an isolated fishing village]. The picture changes completely by the end of the first year, when their Developmental Quotient *sinks*[3] to the astonishingly low level of 72" (**43**, p. 58) and, "By the end of the second year the Developmental Quotient *sinks*[4] to 45 . . ." (**43**, p. 70). Elsewhere he states, "The children . . . [in Foundling Home] though starting at almost as high a level as the best of the others, had spectacularly *deteriorated*"[5] (**43**, p. 59). He maintains that his data show the "comparison of the development in Nursery and Foundling Home" (**43**, p. 67, 69). This emphasis on *change* in score and on *development* would seem to imply that on the average the DQ's of the *same* group of children deteriorated from a score of over 130 during their second month to a DQ of 72 by the end of their first year, and to 45 by the end of the second year.

If the reader has drawn this conclusion, it will become apparent that he has made an error.[6] Spitz says (**43**, p. 59) that at the beginning of his study there were 88 children in Foundling Home below 2½ years of age, and only 45 of these were under 1½ years of age, i.e., there were on the average only 2½ children at each month of age from birth to 18 months and less than 4 at each month of age from 12 to 24 months. Hence, if each case was followed longitudinally, that portion of his graphs covering the first year of development could not have been based on more than half of his subjects, and that portion covering the first six months, the period of most rapid decline in developmental quotient, could not have been based on more than a fifth of his cases. It would be impossible for the graphs to represent the change of performance for a constant sample of 88 children. Rather it appears that his graphs present the average DQ of different children at each age. That this is the correct interpretation becomes eviduring these rounds the chart of each child is inspected as well as the child itself" (**43**, p. 62). It would appear that we may be somewhat skeptical about the general physical environment and care of these infants if they are *actually* confined in holes in their beds so deep that they can not turn in any direction.

---

[3] Italics mine.

[4] Italics mine.

[5] Italics mine.

[6] Maslow and Mittelmann appear to make this error in their summary of Spitz's work (**34**).

Spitz says that "As soon as the babies in Foundling Home are weaned, the modest human contacts which they have had during nursing at the breast stop, and their development falls below normal" (**43**, p. 66). While "The time when the children in Foundling Home are weaned is the beginning of the fourth month" (**43**, p. 66), we note that " . . . the separation from the mother took place beginning after the third month, but *prevalently*[9] in the *sixth*[10] month" (**45**, p. 331).

Let us consider now the time-location of the "decline" in DQ according to his graphs (**43**, p. 69; **47**, p. 272). During the second month the average DQ was approximately 131 for Foundling Home; in the fifth month approximately 88; in the sixth month approximately 76; in the seventh month approximately 75, and at a year approximately 72. Thus we see that there was a difference of approximately 59 DQ points between the children in their second month and the children at a year of age. However, 43 points of this drop occurred by the time the children were tested in their fifth month and hence while the majority of the mothers were still present. The difference between the children in their fifth month and those in their sixth month, the month in which separation from the mother was most prevalent (**45**, p. 331), is 12 points. The difference between the children in their sixth month and the children at a year is only approximately four points.

The data just reviewed support Spitz's contention that marked retardation characterizes the Foundling Home infants; however, the information given by Spitz, rather than supporting his hypothesis that the retardation is due to separation of mother and child, indicates that it was in evidence before the separation.

As Spitz places a great deal of weight on the data derived from the Hetzer-Wolf baby tests, a careful evaluation of them is necessary.

The Tests Used in the Study

As indicated earlier (**43**, p. 72), the most prominent feature of the disorder of "hospitalism" is severe developmental retardation as measured by the Hetzer-Wolf baby tests. It plays an equally important role in the diagnosis of anaclitic depression (**45**). As his reference to these tests he gives Hetzer and Wolf's article "Baby Tests," published in 1928 (**23**). Since he always makes reference to the originally published tests rather than to the revision by Frankl and Wolf (**11**), there can be little doubt that the original scale was used.[11] (These tests are frequently referred to as the "Bühler scales" [**22**, **27**, **32**].)

Inasmuch as the baby tests were originally standardized on Viennese children, the norms may or may not be applicable to children in other cultures. Perhaps some consideration should be given to the original standardization before studies from other cultures are considered.

---

[9] Italics mine.

[10] Italics mine.

[11] The above assumption may or may not be justified inasmuch as he says regarding the tests, "They provide . . . quantifiable data on six distinct sectors of personality . . . " (**43**, p. 55), and only four "sectors of personality" were indicated in the original form of the tests. Six were designated in the revised tests, but the revision was made by Frankl and Wolf and it would appear inaccurate to designate these latter tests as the "Hetzer-Wolf baby tests."

Standardization

According to Hetzer and Wolf (**24**, p. 203), the tests were standardized on 20 children at each month level. The subjects were children in the Reception House for Children in Vienna (**24**, p. 203). Most of the children of the Reception House came out of family care to spend a three-week quarantine period before being placed either with a foster mother or in another institution (**10**, p. 6). Hence these children were themselves subject to maternal deprivation prior to testing. According to Bühler, "Children used in these tests were largely taken from the poorer population of Vienna" (**10**, p. 89), and hence as Herring notes (**22**), the norms are not representative of Viennese infants in general.

Studies on Infants of Other Cultures as Related to Spitz's Findings

The average DQ. In a search of the literature, three studies, all cross-sectional, were found which are relevant to the use of the test with children from other cultures. All three cast doubt on the accuracy of the norms, at least for children of other cultures. The first of these studies was carried out by McGraw in her comparative study of a group of southern white and Negro infants (**32**). "About fifty per cent of the babies [in Tallahassee, Florida] born within the year—both white and colored—figured as subjects in this comparative study" (**32**, p. 28). "Although the subjects were selected at random and the results for both white and colored yield [for the total sample] a normal distribution in terms of 'Developmental Quotient,' the educational levels of the parents tend toward the upper grades, high school, and college levels for both white and colored . . . " (**32**, p. 33–34). McGraw reports the average DQ for each age group on the Hetzer-Wolf test.[12] These are presented in Figure 1. As is apparent in the figure, the average DQ tends to decrease with age for both of McGraw's groups.

The second of these studies was that carried out by Ruth Hubbard on the reliability and validity of the scale (**27**). She does not give average DQ's by month for her 78 babies ranging from 1 to 20 months, but she does conclude from her study that the scale is well worth the efforts of a thorough standardization on American infants. She notes a number of disadvantages of the scale including " . . . the wide scatter of successes necessitating a testing period of an hour in many cases, the hiatus of a month between the eleven- and twelve-month levels for which no tests are provided, the number of frustrations in the series, and the inadequate standardization of administration and scoring" (**27**, p. 382).

The third study was carried out by Hsu with Peiping babies (**26**). While there were only twenty-six infants in his study, the trend of his averages are much like those of McGraw. Hsu concludes, "It is apparent that the Viennese scale rated the Peiping babies too high at the lower age level and too low at the upper age level" (**26**, p. 217).

---

[12] It should be noted that certain changes were made by McGraw in the test material and in scoring; however, she notes that in the English translation Rowena Ripin adopted a scheme of scoring not very different from that employed by Bühler and her associates. In her monograph McGraw notes a number of features of the test which she considers to be " . . . outstanding frailties in the scale as a means of measuring the development of infants" (**32**, p. 65–67).

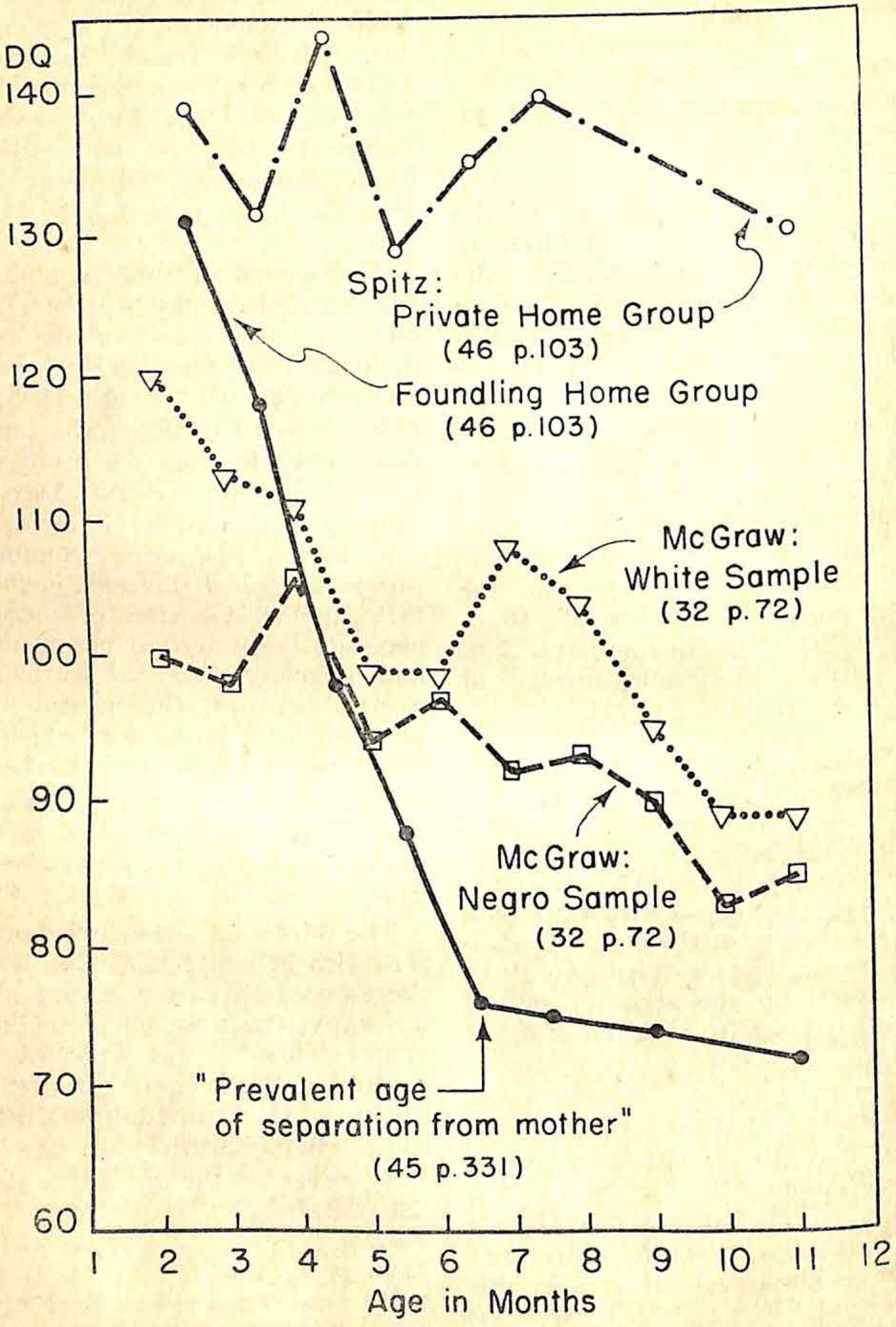


FIG. 1. MEANS OF DEVELOPMENTAL QUOTIENTS WITHIN THE FIRST YEAR FOR THE HETZER-WOLF "BABY TESTS"

Spitz's curve for the Foundling Home infants (**43**, p. 69), is included in Figure 1. The average DQ's of his Foundling Home infants, like the average DQ's in the studies of McGraw and Hsu, decrease with age during the first year. An analysis of this figure supports the contention of Hsu that the Viennese infant scale rates children too high in the early months and too low in the later months. It is regrettable that Spitz was apparently unaware of this problem.[13] However, Spitz's Nursery group did not show a decline in DQ with age. Discussion of the results from this group will be deferred until a later section.

The possibility that the Frankl-Wolf revision of the Hetzer-Wolf baby tests was used in the study should not be overlooked. Two major studies of this scale were found in the literature. The first of these was by Ackerman (**1**) utilizing 25 children at each of four levels (7 months, 8 months, 9 and 10 months, and 11 and 12 months). The average DQ was 106.67. There was only a slight drop in the average DQ for these ages, the drop in means being 2.22 points from the tests at 7 months to those at 11 and 12 months. However, it should also be noted that during this period the differences in DQ of the Foundling Home children in Spitz's study were essentially of the same magnitude, being only three or four points lower at 12 months than they had been at 6 to 7 months. The study sheds no light on those months of most concern to us, those in which Spitz's age differences were "significant." In contrast, we may consider Herring's study (**22**) on the reliability of the test. In her report on 114 subjects the mean DQ is given for each group. The same trend as was noted for the original scale (i.e., for the averages to be successively lower) is present, although it is not as marked. (The mean DQ's for 1, 2, 3, 5, 6, and 7 months were respectively 132.7, 128.1, 90.9, 115.5, 119.8, and 107.6.) This trend suggests that the scale, even when revised, is inadequately standardized and gives a spuriously high score at the earlier ages.

Spitz also used the Hetzer-Wolf baby tests on a group of 17 children raised in private homes of white-collar workers. From a close consideration of these homes, he felt that the relationship between the child and his mother was exceptionally good (**46**, p. 94). The average DQ's for these children for their first year were taken from his chart (**46**, p. 103) and are included in Fig. 1. One cannot determine from his chart or from the text of his article whether or not this curve is based on longitudinal data, nor can one tell how many of these children are represented at any one age.

It should be noted that Herring found no significant difference between the scores of infants from the three upper and three lower socio-economic groups on the Frankl-Wolf tests (**22**). In essence, being from "private homes of white collar workers" does not predispose children of this age to have high scores. It is not apparent why Spitz's findings on these children reared in private homes are so out of line with Herring's report (**22**), with the findings of Bayley (**7**) that scores during early infancy have little relation to scores in the latter part of infancy, and with the findings of Bayley and Jones (**8**) that scores within the first year are essentially unrelated to parents' socioeconomic, educational, and occupational level.

---

[13] Bühler maintains that in all studies called to her attention the median DQ has not differed significantly from a DQ of 100 (**11**, p. 88). If one considers the samples as a whole, without regard to age-level, the preceding studies would probably also agree with this contention; however, this would not raise the validity of the instrument as a measure of normal development at any given age level.

From the studies summarized in this section we may conclude that with most groups which have been tested, the Hetzer-Wolf tests give average DQ's which are considerably above 100 in the earlier months and considerably below in the later months. The Frankl-Wolf tests also appear to give average DQ's which are considerably above 100 in the early months. These considerations suggest that a considerable portion of the difference between the scores of the younger and older infants of Foundling Home could have been predicted on the basis of normative studies, and that these differences may be chiefly an artifact of the tests' standardization. Spitz's data obtained from his other groups are out of line with the majority of studies of the Hetzer-Wolf tests.

Predictive Ability of the DQ

Although the extent of the relationship between tests at different ages has not been determined for the Hetzer-Wolf baby tests, some information is available for the Frankl-Wolf revision. Herring (**22**) correlated the average of her subjects' scores on two successive tests at one month with their average on tests at five and six months and at nine and ten months. The correlations were respectively .288 for 22 subjects and .345 for 23 subjects. Neither of these is significant at the 5 per cent level. Her correlation of the average of scores at five and six months with the average of scores at nine and ten months for 22 cases was .448. While this correlation is significant at the 5 per cent level, it is still too small to be of much value for predictive purposes.

The preceding study indicates that one would do little better than chance in predicting from a single administration of the Frankl-Wolf test whether an infant's score four or five months later will be low or high. Since this test is a revision of the test used by Spitz, it would seem unlikely that the tests he used would have any greater predictive value. Support for this hypothesis, at least for longer time intervals, is found in Anderson's use of most of these items in his study of the predictive value of infancy tests (**3**). Other studies indicate that lack of predictiveness is a characteristic of infant scales; e.g., Bayley (**7**) in a complete analysis of the predictive value of an infant test, the California First Year Mental Scale (**4**), found that scores at six months of age are negatively related to performance at school age and beyond, and that the median correlation between performance at a year and tests subsequent to school age is only .23.

Spitz's reports would lead one to believe that a drop in score within the first year has some intrinsic value and that a low score by the age of one is of special significance. That a drop in score within the first year has little predictive value may be illustrated by considering those subjects in the Berkeley Growth Study whose scores dropped from their test at three months to the one at 12 months. The 21 subjects whose scores dropped during this interval of time had a mean Deviation IQ[14] of 123.5 at age 10 years, which is not significantly different from the mean Deviation IQ of the total Berkeley Growth Study group. Neither does low score at a year have any special significance, as can be shown by determining the mean score at ten years of those BGS subjects whose Deviation

---

[14] The Deviation IQ is defined as the standard score, given a mean of 100 and a standard deviation of 16 (**37**).

IQ's were below 80 at 12 months of age. The mean Deviation IQ for these four subjects at ten years was 131.7.

If the Hetzer-Wolf tests have no greater predictive value than other infant tests (as such evidence as is available indicates), the conclusion is obvious: Even if these infant's scores were below average we would not need to be concerned, because their scores at one year will be essentially unrelated to their performance by the time they have reached school age. This is, of course, in direct contradiction to the position taken by Spitz, for he maintains that the effects of the maternal deprivation on the Foundling Home children has resulted in irreparable damage (**43**, **44**), and that such damage is reflected in the infants' test scores.

The Relation of "Deterioration" to Separation in Cases of Anaclitic Depression

The infants who developed the syndrome of anaclitic depression in the Nursery group were separated from their mothers for a practically unbroken period of three months[15] (**45**, p. 319). Spitz states that "This removal took place for unavoidable external reasons" (**45**, p. 319). It is regrettable that he does not specify these reasons so that the infants' subsequent development can be appraised in the light of them. He reports that 45 of the 95 infants (**45**, p. 336) separated from their mothers show some manifestation of this syndrome (**45**, p. 318) and that "*No* child developed the syndrome in question whose mother was *not* removed" (**45**, p. 320). According to his graph (**45**, p. 329), the average age at which mother and child were separated was seven months. The last test before the separation was at six months and one day and the first test afterwards at seven months and twenty-six days. Between these two dates there was a decline in score from 121.5 to 98. Of this 23.5 point decline there was only a 6.5-point drop in score from the first test after separation to the last test before reunion. Subsequent to reunion there was a 25-point rise. Considerations in the next section would appear to raise questions as to whether these changes are necessarily the result of separation from and reunion with the mother.

TABLE 1

Average Developmental Quotients of Spitz's Nursery Subjects: A Comparison of Average DQ's of Subsamples with Each Other and with the Total Group*

Age (in mos.)	Original Group (**43**, p. 71) $N=69$	Anaclitic Depression Group† (**45**, p. 329) $N=45$	Residual Sample $N=56$	Total Group‡ (**46**, p. 87, 101–103) $N=170$
2–3	98.0			116.5
3–4	105.0			119.0
4–5	110.0			110.5
5–6	110.0			105.0
6.0	*105.0*	121.5	90.2	*104.5*
6–7	100.0			104.0
7–8	110.0			109.5
7.9	*110.0*	98.0	116.6	*109.0*
8–10	110.0	91.5	117.3	107.5
10–12	98.0	116.0	82.0	97.5

* Italicized values were interpolated from Spitz's curves.

† These children were tested at 4 ages: 6 mos. 1 day, 7 mos. 26 days, 9 mos. 15 days, and 11 mos. 0 days.

‡ The total group consisted of 196 cases; however, Spitz's report is based on only 170 of these cases, as the other 26 had not reached the age at which autoreotic activities begin (**46**, p. 87).

In his article on autoeroticism, Spitz presents a graph apparently based on 170 of the 196 infants thus far observed in Nursery (**46**, p. 87,

---

[15] While Spitz maintains that race does ". . . not appear to exert demonstrable influence on the incidence of the syndrome" (**45**, p. 318), it should be noted that severe depression occurred among only 9 per cent of the white children whereas it occurred in 26 per cent of the colored children.

101–103). (The reason for excluding 26 of the cases was that these infants had not yet reached the age at which autoerotic activities usually begin.) Average DQ's for this group of 170 infants were taken from his graph (**46**, p. 87, 101–103) and are presented here in Table 1. This total group can be divided into three subgroups: (*a*) 69 infants in the original group. Average DQ's for this subgroup were obtained from the curve in his first article (**43**, p. 71) and are presented in Table 1. (*b*) 45 infants separated from their mothers and manifesting anaclitic depression. The graph for this group Spitz presents in his article on anaclitic depression (**45**, p. 329). (As he did not indicate on how many cases the graph was based or how many were represented at each age, it can only be assumed that it was based on the 45 cases.[16]) Average DQ's for these infants are given in Table 1. (*c*) Remaining 56 cases of the 170 not accounted for by the preceding two subgroups. From the scores in Table 1 for the total group and for the other two subsamples, the scores for this residual sample of 56 cases were determined for the last six months of the first year and are also included in the table.[17] The fluctuation of the average for these 56 children is marked.

The changes for this residual sample occur at the same ages as for the subjects manifesting anaclitic depression. That is, at the same age that Spitz reports a 29-point drop in score for subjects manifesting anaclitic depression as a result of their being separated from their mothers, this other subsample shows a 27-point rise in score, and at the same age that he reports a 25-point increase in score as a result of the subjects being reunited with their mothers, this other group shows a 35-point drop in score.

As was indicated earlier, it appears that 95 of the children were separated from their mothers and only 45 developed the disorder of anaclitic depression. Spitz does not give the average scores for the other 50 subjects, but if we assume that they are among the 56 subjects of the subsample referred to above, we can determine the average score for the combined group of 101 subjects. The means for the same ages are 104, 108, 106, and 97, respectively. One wonders if the reason for the low scores in the anaclitic depression group is that he assigned to it only those of the 95 cases who showed a marked drop in score. It is impossible to determine from the data he presents whether or not this surmise is correct. However, his contention that the 45 cases manifesting anaclitic depression show a drop and increase in score as a result of separation from and reunion with their mothers would appear to exact little confidence in the light of two other considerations: namely that the other and even larger subsample referred to above shows even more marked fluctuations in score, and the pattern of means which he gives for the total sample of Nursery is completely out of line with those which McGraw (**32**) found for unselected samples of white and Negro infants (cf. Fig. 1).

---

[16] This graph may be based on more than 45 cases as he labels it "Variations of Development Quotient (Average) Under the Influence of Separation from and Reunion with Mother" (**45**, p. 329), and more than this number were separated from their mothers. (According to Spitz, not all children who were separated from their mothers developed this syndrome [**45**, p. 320].)

[17] In order to obtain the scores for the residual sample, at each age at which an average score was given for the anaclitic depression group the average score for the total group and for each subgroup was multiplied by the number of cases in the respective samples. At each age the difference between the product for the total sample and the sum of the products for the subsample was divided by the number of subjects in the residual group to obtain its mean score. This procedure assumes the same number of cases at each age, i.e., it weights the mean value according to the number of cases reported in the sample.

Physical Status of the Children

Spitz gives certain information concerning the physical condition of the infants. With regard to the children in Nursery, he says,

> During the *three-and-a-half years*[18] of our study of Nursery, we had occasion to follow 122 infants, each for approximately a full year. During this time *not a single child died.* The institution was visited by no epidemic. Intercurrent sickness was limited, on the whole, to seasonal colds, which in a moderate number developed into mild respiratory involvement; there was comparatively little intestinal disturbance; the most disturbing illness was eczema. The *unusually high*[19] level of health maintained in Nursery impelled us to look into its past record (**44**, p. 117).

This almost glowing picture of health on 122 infants followed for approximately a full year receives a somewhat different cast when in the second article in this same volume he discusses the presence of anaclitic depression in 45 (**45**, p. 320) of these infants. He says, "Nor can the infant enact a suicide; but it is striking that these cases *one and all*[20] show a great susceptibility to intercurrent sickness" (**45**, p. 320). Forty-five children constitute 37 per cent of the total sample. If such great susceptibility to "intercurrent sickness" is shown in 37 per cent of the children, it would seem difficult to maintain without qualification that, on the average, an "unusually high level of health [was] maintained in Nursery."

However, the physical condition of the children of Nursery needs less attention than that of the Foundling Home children. According to Spitz,

> The children in Foundling Home showed all the manifestations of hospitalism, both physical and mental. In spite of the fact that hygiene and precautions against contagion were impeccable, the children showed, from the *third*[21] month on, extreme susceptibility to infection and illness of any kind. . . . No figures could be elicited on general mortality; but during my stay an epidemic of measles swept the institution, with staggeringly high mortality figures, notwithstanding liberal administration of convalescent serum and globulins, as well as excellent hygienic conditions. Of a total of 88 children up to the age of 2½, 23 died. . . .
>
> In view of the damage sustained in all personality sectors of the children during their stay in this institution, we believe it licit to assume that their vitality (whatever that may be), their resistance to disease, was also progressively sapped (**43**, p. 59).

This quotation can better be appreciated when it is recalled that this was a cross-sectional study at the time this was written, and that his initial study of this group was apparently done in a short period of time, perhaps in less than a week. Thus it appears that the study well may have been carried on at a time when the children were ill. If so, it might be expected that the children would appear to have a lowered vitality. However, one would also think that Spitz would have been surprised at how well the children performed on the Hetzer-Wolf tests, rather than at how "poorly" they did.

Although he notes that from the third month the children showed extreme susceptibility to infection and illness of any kind, he could only assume that their vitality, their resistance to disease, was progressively sapped during their stay in the hospital, as he did not observe the children except for an extremely brief period of time. From this and his subsequent articles (**44**, **45**) it is clear that the purported cause is psychosomatic damage resulting from unfavorable environmental conditions. Spitz feels that the most important of these environmental conditions is the separation of the child from his mother. Such a conclusion would hardly seem warranted in regard to health in the fourth and fifth months since " . . . the separation from the mother took place beginning after the third month, but prevalently in the sixth month"[22] (**45**, p. 331).

---

[18] Italics mine.

[19] Italics mine.

[20] Italics mine.

[21] Italics mine.

It appears that Spitz's comparisons of these infants before and after separation from the mother warrant little credence, as they were based on cross-sectional data. A less conservative evaluation would maintain that his conclusions must be rejected as he was contrasting two groups of children of whose former and future development he was not cognizant.

To obtain the flavor of Spitz's "longitudinal study" of these infants, the following paragraph is quoted from his article published in 1951; "The progressive deterioration and the increased infection-liability lead in a distressingly high percentage of these children *to marasmus and death*.[23] Of the 91 children followed by us for two years in Foundling Home, 37 per cent died . . . " (**47**, p. 271). If the reader assumes that 91 children were followed for two years, he is making an erroneous assumption, because 23 of these children (25 per cent) died of measles during the short time that Spitz was initially there (**43**, p. 59), another four died before the end of the first year, and 7 died in the second year; hence, it would have been possible to follow only 57 of these children. However, according to his own report, 36 more of these could not be learned about in the follow-up study (**44**, p. 114), 23 because they had been taken back to their families, 9 because they had been adopted or placed in other institutions, and 4 could not be accounted for. At one point the number followed for two years is clearly stated by Spitz (**44**, p. 114) to have been 21. It would appear hazardous to base conclusions on such a small portion of the original sample, inasmuch as the selective factors operating in its reduction are unknown.

The preceding quotation from Spitz might lead one to believe that marasmus was the most frequent cause of death; however, it should be noted that this "disease" was not even mentioned in his discussion of the death of these 34 infants in either of his first two reports. In the second of these he says, "In the course of the first year, 27 of these died of *various*[24] causes, among which were an epidemic of measles, intercurrent sickness, and cachexia; by the end of the second year, another 7 of those originally seen had died; this represents a total mortality of over 37 per cent in a period of two years" (**44**, p. 114). As noted above, measles was the most frequent of these various causes of death. It seems surprising that Spitz neglected to mention marasmus both in his first report (**43**) and in his second report (**44**) since it was present in such a "distressingly high percentage of these children."[25]

---

[22] In contrast to this group, breast-feeding in the Nursery group was less frequent; thus Spitz says, "In Nursery this percentage is smaller, so that in most cases a formula is soon added, and in many cases weaning takes place early" (**43**, p. 61).

[23] Italics mine.

[24] Italics mine.

[25] While in the second article Spitz states that cachexia was one of the causes of death, it should be noted that this disorder was *not* mentioned in the first article on "hospitalism" (**43**). He does not specify which of the more than a dozen different kinds of cachexia this is,

Regarding these infants' physical development, Spitz gives only cursory information. In his follow-up report on Foundling Home infants, he states that only two have attained normal height of the two-year-old, and that only three of the subjects "fall in the weight range" of normal two-year-olds. This retardation amounts to more than a year in a number of the cases and is indeed very marked; it is especially marked in height, and we are left with the inference that maternal separation is responsible for this retarded growth. It would seem that the prediction may then be made that children separated from their mothers as in Foundling Home will be drastically retarded in growth and height. As we have not usually believed that skeletal growth is to so great a degree responsive to psychological influences, alternative interpretations should be considered: (*a*) separation from the mother is the cause of extreme retardation in physical growth; (*b*) this sample (after removal of the 34 cases who died and of the 36 who were not available for the follow-up study) constitutes an inferior biological selection; (*c*) the norms used by Spitz (not stated) are not applicable to this population (not described).

The first alternative appears to claim more, especially with regard to physical growth, than students of child development are likely to accept without confirming evidence from more carefully controlled studies. The second and third alternatives, apparently discarded by Spitz, must be regarded as reasonable unless negated by further data.

SUMMARY

From a study of several groups of infants from contrasting environments, Spitz concludes that infants, as a consequence of being separated from their mothers, develop psychological disorders. These disorders are supposedly manifest in both "mental and physical" symptoms. A careful consideration of these studies indicates that they were so planned and carried out that they could give neither positive nor negative evidence for his hypotheses.

Some of the major vulnerabilities in Spitz's report of his studies may be summarized as follows: He fails to indicate the dates and places of the studies and neglects to indicate the composition and training of the research staff. He is inconsistent in his report of the number of children present in his studies, and his descriptions of their parents are contradictory. The groups which are compared as to mental and emotional development apparently differ in racial extraction, socioeconomic background, and in heredity. He is inconsistent in his descriptions of their physical surroundings, of their care, and of their physical health. While most of his conclusions could only be considered plausible if based on longitudinal data, most of them are based on cross-sectional data, or at best on a mixture of the two. His reports regarding the average amount of time spent in observing the children are inconsistent and other considerations indicate that there was a great deal of variability in the amount of time different infants were observed. He uses tests with means which apparently decrease with age. He assumes that test results at one age in infancy are highly related to test results several months later and that the same numerical quotient at different ages has the same relative significance, assumptions which are unwarranted in terms of our present state of knowledge.

The preceding conclusions make it

---

and as far as the writer has been able to ascertain, no one of the forms is synonymous with marasmus.

clear that the results of Spitz's studies cannot be accepted as scientific evidence supporting the hypothesis that institutional infants develop psychological disorders as a result of being separated from their mothers.

CONCLUSION

The writer does not doubt the potential advantages of maternal as compared with institutional care. An analysis of institutional care, involving procedures in a large number of institutions, would be expected to show a somewhat different pattern of benefits and of hazards than in the case of children brought up in their own homes. As yet, however, we do not have convincing evidence, based on scientifically controlled investigations, as to any of the major problems in this area.

THE FISCHER STUDY OF HOSPITALISM[26]

Fischer (**16**) comes to conclusions similar to Spitz's in a study of infants in a Catholic home for unmarried mothers, the St. Agnes Home in West Hartford, Connecticut. The subjects used in this study consisted of 62 out of 189 infants tested between the chronological ages of six and seven months in the years 1946 to 1949. The children were tested with the Cattell infant test and those who had IQ's below 90 were chosen for the study.

Fischer maintains that the proportion of children falling below an IQ of 90 exceeds expectancy (62 of 189 children, or 33 per cent). She gives no basis for this contention and it seems highly questionable. If we assume an *SD* of 16 for IQ's on this scale at the age of six months, we would expect 26.5 per cent of a normal sample of subjects to fall below an IQ score of 90. The difference between this proportion and Fischer's is not significant. (As far as the writer knows, *SD*'s for each month for this scale are not available.)

Fischer considers that the low scores for these infants are the result of their being institutionalized; thus she says, "They most probably are but an indication that the infant in question has reacted early and typically to institutional frustation . . ." (**16**, p. 531). This contention is obviously unacceptable inasmuch as she chose those children with IQ's below 90 and the number of cases which she found is not significantly greater than one would expect, assuming an *SD* of 16 points. In fact, if on an intelligence test, one did not find a sizable proportion of the cases falling below 5/8 standard deviations at each age, either the test would be suspect or one would suspect a bias in the selection of cases.

The mean IQ for this group of subjects was 76.11. Assuming a chronological age of 6.5 months, this would mean an average mental age of 4.9 months, i.e., the children were on the average 1.6 months retarded. In discussing these children Fischer says,

> So far our six-month-old, suffering from what we consider a form of hospitalism, emerges therefore as a baby with good sensory, social and muscular strength reaction who, however, does not engage in the grasping activities which to an essential degree characterize the normal six-month-old (**16**, p. 528).

From the preceding quotation and description of the subjects, it is apparent that Fischer places a great deal of weight on the child's performance on the Cattell infant scale. It seems almost as if she would maintain that an infant in an institution who makes an IQ score below 90 is afflicted with "hospitalism." However, it is apparent that if one selected children at the age of six months who are retarded to the extent of this sample, that is, children who on the average are performing only at the level of the typical child 4.9 months old, they could not be expected to perform the grasping activities of the average child of six months of age. Hence we cannot accept her conclusion, on the basis of this sample, that lack of this ability is a characteristic of "hospitalism."

---

[26] The quotations in this section from L. Fischer, *Amer. J. Orthopsychiat.*, 1952, **22**, 522–533, are made with the permission of the American Orthopsychiatric Association, Inc.

Fischer seems to feel that the children subsequently placed in adoptive homes improve in performance on the scale as a consequence of the placement. She states, "Even including these unfavorable scores, the mean IQ of the adopted children (after already increasing in the institution to 86.23) is as high as 97.54. This becomes particularly interesting if we compare the mean scores of all the children in their development" (**16**, p. 525–526). The mean IQ of the 36 children who were re-examined in their adoptive homes at an average age of 20 months was 97.54. If the tests at these three ages were completely unrelated, one would expect the scores at the subsequent ages to be normally distributed around the mean of the population *which they represent*. However, a slight relationship may be present. For the purposes of this article and for the population which this group represents, the author will assume a mean of 100 at all three ages, while at the same time recognizing that the selective factors in operation make this assumption very vulnerable. If standard deviations were available for the Cattell test for the age levels 6, 11, and 20 months, and if correlations between tests at these age levels were available, scores could be predicted from the initial test at six months to the subsequent tests at 11 and 20 months. As far as the author knows, such statistical material is not available. Hence, predictions will be made on the basis of material reported in other studies. Continuing to assume an *SD* of 16 at all three ages, and on the basis of Bayley's work (**7**) assuming a correlation of .52 for tests given at age 6 and 11 months, a correlation of .23 for tests given at 6 and 20 months, and a correlation of .60 for tests given at 11 and 20 months, these scores will be predicted. The predicted scores together with the standard error of estimate are as follows: For 11 months, 87.6 ± 13.67 and for 20 months, 94.5 ± 15.57 and 91.7 ± 12.80. Granted the previous assumptions, it is apparent that the values obtained by Fischer are not significantly different from those expected purely on the basis of regression toward the mean.

After selecting children retarded in development, children with a mean IQ of 76.1, she states when she comes to a description of their total behavior: "Obviously, the children we observed are youngsters of average potentialities, whose atypical reaction to a test situation is environmentally fostered" (**16**, p. 529). Since a number of studies have shown that there is very little relationship between tests during infancy and future performance, the test results would not provide a basis for questioning the statement that these subjects "are youngsters of average potentialities." On the other hand the validity of this contention is far from *obvious*, and no information is included in the text of the article which makes it more than plausible. It would seem that such an assumption would have to be based on adequate information about both parents. A negative selective factor is suggested in the "sexual indiscretion" of the mothers and perhaps by their coming to a maternity home for unmarried mothers. The writer was unable to find in her article any support for the contention that their "atypical reaction" to the test was environmentally fostered.

Fischer subdivides her group of six- to seven-month-old subjects into two other groups. She states,

> Again patterning is very clear-cut, as 87 per cent of the children, equally divided, comprise two dominant groups. One of them is essentially passive in all areas—including social and sensory areas. These are children who in the test performance do not react to sound or to the mirror. . . . The other pattern is that of marked social responsiveness—from friendly smiles for the examiner or their own picture in the mirror to tremendous demandingness (**16**, p. 529).

The two dominant groups would be composed of 23 children each. Fischer states that the first group of 23 children do not react to sound or to the mirror. However, it is difficult to reconcile this statement with her earlier ones: that 94 per cent of the 53 subjects turn to the sound of a voice, and that "Significant changes, however, occur in all but two items at the five-month level, with an insignificant drop on item 1 (turning to the sound of a bell)" (**16**, p. 528). Ninety-four per cent of the 53 subjects would include 50 of the infants. If we assume that all "nonreactors" to sound were in this group, 20 of the 23 children would still have reacted to sound.

Fischer does not give a sufficient description of the second group's performance on the test to permit an adequate evaluation; however, from the description given of both groups, the question is raised in the author's mind as to whether the same subgroups could have been obtained merely by dividing the children with IQ's below 90 into two groups, those most and least retarded.

The preceding considerations would seem to raise considerable doubt that Fischer's investigation supports Spitz's conception of hospitalism.

REFERENCES

1. Ackerman, Dorothy S. The critical evaluation of the Viennese Tests applied to 200 New York infants six to twelve months old. *Child. Develpm.* 1942, **13**, 41–53.
2. Anastasi, Anne, & Foley, J. P. *Differential psychology.* (Rev. Ed.) New York: Macmillan, 1949.
3. Anderson, L. D. The predictive value of infancy tests in relation to intelligence at five years. *Child Develpm.*, 1939, **10**, 203–212.
4. Bayley, Nancy. *The California First-Year Mental Scale.* Berkeley: Univer. of California Press, 1933.
5. Bayley, Nancy. Mental growth during the first three years: A developmental study of sixty-one children by repeated tests. *Genet. Psychol. Monogr.*, 1933, 14(1), 1–92.
6. Bayley, Nancy. *The California Infant Scale of Motor Development.* Berkeley: Univer. of California Press, 1936.
7. Bayley, Nancy. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, **75**, 165–196.
8. Bayley, Nancy, & Jones, H. E. Environmental correlates of mental and motor development; a cumulative study from infancy to six years. *Child Develpm.*, 1937, **8**, 329–341.
9. Bowlby, J. Maternal care and mental health. *Bull. World Health Organ.*, 1951, **3**, 355–534.
10. Bühler, Charlotte, & Hetzer, Hildegard. *The first year of life.* New York: John Day, 1930.
11. Bühler, Charlotte, & Hetzer, Hildegard (Eds.). *Testing children's development from birth to school age.* New York: Farrar and Rinehart, 1935.
12. Cattell, Psyche. *The measurement of intelligence of infants and young children.* New York: Psychological Corporation, 1947.
13. Dennis, W. The effect of restricted practice upon the reaching, sitting, and standing of two infants. *J. genet. Psychol.*, 1935, **47**, 17–32.

14. DENNIS, W. Developmental theories. In W. Dennis (Ed.), *Current trends in psychological theory*. Pittsburgh: Univer. of Pittsburgh Press, 1951. Pp. 1–20.
15. DENNIS, W., & DENNIS, MARSENA G. Behavioral development in the first year as shown by forty biographies. *Psychol. Rec.*, 1937, **1**, 349–361.
16. FISCHER, LISELOTTE K. Hospitalism in six-month-old infants. *Amer. J. Orthopsychiat.*, 1952, **22**, 522–533.
17. FRANKL, LISELOTTE, & WOLF, KAETHE. The test for the first year of life. In Charlotte Bühler and Hildegard Hetzer (Eds.), *Testing children's development from birth to school age.* New York: Farrar and Rinehart, 1935. Pp. 95–127.
18. GELEERD, ELIZABETH R. A contribution to the problem of psychoses in children. *Psychoanal. Stud. Child.* **2**, 271–291. New York: International Univer. Press, 1946.
19. GESELL, A. L. *Infancy and human growth.* New York: Macmillan, 1928.
20. Halpern, L. J. *How to raise a healthy baby.* New York: Prentice-Hall, 1940.
21. HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. M. Comments on the formation of psychic structure. *Psychoanal. Stud. Child*, **2**, 11–38. New York: International Univer. Press, 1946.
22. HERRING, AMANDA. An experimental study of the reliability of the Bühler Baby Tests. *J. exp. Educ.*, 1937, **6**, 147–160.
23. HETZER, HILDEGARD, & WOLF, KÄTHE. Baby tests. *Z. Psychol.*, 1928, 107.
24. HETZER, HILDEGARD, & WOLF, KÄTHE. Baby tests. In Charlotte Bühler and Hildegard Hetzer (Eds.), *The first year of life.* New York: John Day, 1930. Pp. 189–239.
25. HILGARD, ERNEST R. *Introduction to psychology.* New York: Harcourt Brace, 1953.
26. HSU, E. H. On the application of Viennese Scale to Peiping babies. *J. genet. Psychol.*, 1946, **69**, 217–220.
27. HUBBARD, RUTH M. A study of the reliability and validity of the Bühler infant scale. *J. genet. Psychol.*, 1935, **47**, 361–384.
28. JONES, H. E. Homogamy in intellectual abilities. *Amer. J. Sociol.*, 1929, **35**, 369–382.
29. JONES, H. E. The environment and mental development. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd Ed.) New York: Wiley , 1954.
30. KRIS, E. Notes on the development and on some current problems of psychoanalytic child psychology. *Psychoanal. Stud. Child*, **5**, 24–46. New York: International Univer. Press, 1950.
31. LOEVINGER, JANE. On the proportional contributions of differences in nature and in nurture to differences in intelligence. *Psychol. Bull.*, 1943, **40**, 725–756.
32. MCGRAW, MYRTLE B. Comparative study of a group of Southern White and Negro infants. *Genet. Psychol. Monogr.*, 1931, **10**(1), 1–105.
33. MCNEMAR, Q. *The revision of the Stanford-Binet scale, an analysis of the standardization data.* New York: Houghton Mifflin, 1942.
34. MASLOW, A. H., & MITTELMANN, BÉLA. *Principles of abnormal psychology.* (Rev. Ed.) New York: Harper, 1951.
35. ORLANSKY, H. Infant care and personality. *Psychol. Bull.*, 1949, **46**, 1–48.
36. PINNEAU, S. R. A critique on the articles by Margaret Ribble. *Child Develpm.*, 1950, **21**, 203–228.
37. PINNEAU, S. R. Changes in performance on intelligence tests from one month to eighteen years. Paper read at West. Psychol. Ass., Seattle, June, 1953.
38. RIBBLE, MARGARET. *The rights of infants.* New York: Columbia Univer. Press, 1943.
39. RIBBLE, MARGARET A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 2. New York: Ronald Press, 1944. Pp. 621–651.
40. RUCH, FLOYD L. *Psychology and life.* (3rd Ed.) Chicago: Scott, Foresman, 1948.
41. RUCH, FLOYD L. *Psychology and life.* (4th Ed.) Chicago: Scott, Foresman, 1953.
42. SHIRLEY, MARY M. *The first two years: Vol. I. Postural and locomotor development.* Minneapolis: Univer. of Minnesota Press, 1931.
43. SPITZ, R. A. Hospitalism. An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child*, **1**, 53–74. New York: International Univer. Press, 1945.
44. SPITZ, R. A. Hospitalism: A follow-up report. *Psychoanal. Stud. Child*, **2**, 113–117. New York: International Univer. Press, 1946.
45. SPITZ, R. A. Anaclitic depression. *Psychoanal. Stud. Child*, **2**, 313–342. New York: International Univer. Press, 1946.
46. SPITZ, R. A., & WOLF, KATHERINE M.

Autoerotism. Some empirical findings and hypotheses on three of its manifestations in the first year of life. *Psychoanal. Stud. Child*, **3–4**, 85–120. New York: International Univer. Press, 1949.

47. Spitz, R. A. The psychogenic diseases in infancy: An attempt at their etiologic classification. *Psychoanal. Stud. Child*, **6**, 255–275. New York: International Univer. Press, 1951.

48. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* New York: Houghton Mifflin, 1937.

49. Wolff, G. Maternal and infant mortality in 1944. *U. S. Children's Bur. Stat. Series No. 1*, Fed. Sec. Agency, Soc. Sec. Admin., U. S. Children's Bureau.

50. Milk—and love. *Time*, 1952, **59**, (18), 51.

Received July 7, 1954.

PSYCHOLOGICAL BULLETIN
Vol. 52, No. 5, 1955

REPLY TO DR. PINNEAU

RENÉ A. SPITZ
New York City

Pinneau has expended an extraordinary amount of sagacity and labor in an attempt to discredit every facet of the research I have published. To reply to each of his innumerable points; to correct his misunderstandings, arbitrary conclusions, and unwarranted assumptions would require more space than his article. I will therefore limit my answer to only a few of the most confusing major arguments advanced.

My research, conducted for close to twenty years and comprising the intensive and systematic observation of over 400 infants in their first year, was reported in three or four dozen scientific and professional articles, as well as in about a dozen motion pictures. Of these Pinneau has picked out five articles for the purposes of his polemic. If we disregard the numerous less essential points contained in his paper, we find that he challenges primarily the three following conclusions to which I had arrived in my research:

1. That affective interchange is paramount, not only for the development of emotion itself in infants, but also for the maturation and the development of the child, both physical and behavioral.
2. That this affective interchange is provided by the reciprocity between the mother (or her substitute) and the child.
3. That depriving the child of this interchange is a serious, and in extreme cases, a dangerous handicap for its development in every sector of the personality.

Pinneau endeavors to invalidate these conclusions with the help of statistical manipulations on one hand, and on the other by implying that the experimental psychological material I presented is unsound. To begin with, I wish to make it clear that the experimental psychological and statistical material in the five articles discussed by Pinneau was not introduced by me to *prove* my point, as the articles are addressed to medical readers. They were used as supportive evidence, subordinated to the clinical data—an illustration, as it were, of the description presented. Dismissing these statistics as inadequate would therefore not invalidate my clinical findings. It will, however, be shown further on that the statistics presented by me are correct, and therefore independently confirm and support the clinical findings.

To achieve his purpose, Pinneau has laid a sort of patchwork of numbers. His numerical speculations rest on figures culled, in a biased manner, from five of my publications; these five publications, widely separated in time, deal with different phases of my research. To these mathematics he added various unsupported claims which are in contradiction with the actual facts. With their help he tries to prove that the observations on which my findings and conclusions are based are not longitudinal, but cross-sectional. Once it is proved that they are cross-sectional, he argues, then the statements advanced by me are invalid. Pinneau has not defined what he means by longitudinal. For the purposes of a study on the first year of life longitudinal has to be defined as comprising a period sufficient to detect significant developmental changes in the subject. In the first year of life such a study will require at least two and preferably three months.

I will begin with Pinneau's treatment of the cases in the original Nursery group. He has confused the figures of the original sample of 69 children observed in Nursery for the purposes of the study on "Hospitalism" (**16**, **17**) with the sample of 123 children observed in the same institution for the purposes of "Anaclitic Depression" (**18**). He posits an arbitrary date for the beginning of observations in Nursery and thus succeeds in creating the impression that only one-third or at best one-half of the children in question in either sample were observed consistently for a full year. He concludes from this that "the graphs based on the sample of children must to a large extent represent different children at each age," a wording which creates in the reader's mind the impression that most of the children were seen only once or twice in the course of their first year.

I can dispel the confusion created by Pinneau. The 69 children who form the sample of the article on "Hospitalism" (**16**) consist of:

39 children seen for 12 consecutive months or more from birth (within two weeks of delivery);
15 children seen for 11 months or more, beginning within the first month of life;
9 children seen for 11 months, beginning within the first month of life;
6 children seen for 10 months, beginning within the second month of life.

Let me add for those not familiar with infant observation that no test of any kind has yet been devised to provide really useful psychological information on the infant in its first two months of life, during which behavior is more in the nature of reflex than response. Therefore we did not consider it particularly important that a minority of our sample was not seen by us in the first month and a few of them not in the second month. It might also be mentioned that when Pinneau discusses the usefulness of the Viennese test, most of the discrepancies which he makes so much of are caused just by the unreliability of all and any test scores in the first two months of life. Each of the 69 children was seen at least once a week during the whole period of his stay by myself and/or my associates. It seems to me that the 30 days by which some of the children fall short of the full year's observation hardly can detract from the longitudinal character of the study. When possible each child was tested once a month. When this was not feasible, bimonthly intervals were observed.

After having put in question the original sample of 69 children dealt with in "Hospitalism" (**16**), Pinneau casts doubts upon the procedures used with subsequent children in the Nursery. The fact is that the populations that were discussed in the studies in my subsequent articles ("Anaclitic Depression" (**18**), "Autoerotism" (**21**), "Psychogenic Diseases" (**23**)) were routinely observed beginning with the first day when the child arrived (that is, within two weeks of birth) at Nursery. A film record was made of its first responses, and the observations continued until the child left the institution, that is, after the completion of the first year.

Pinneau's argumentation on the quantitative aspects is therefore invalid when all the facts and not merely a selected portion of them are examined.

We will now proceed to examine other of Pinneau's treatments of data "to obtain the flavor" (as he puts it) of his criticism.

Perhaps this is the place to mention that a good many of Pinneau's interpretations were made possible by the special circumstances surrounding my research. The physician has a professional obligation of discretion to his patients. Furthermore, penal institutions are forbidden by law to give any clue to the identity of their inmates and the same goes for Foundling Home. I made every effort to keep identity, time, and place protected from uncalled-for inquisitiveness, to the point of using, here and there, misleading clues, such as the term "Western Hemisphere." For the reader's information that term was used by me in a cultural sense, meaning the Western world, including Europe.

When therefore I avoided publishing any specific date on the sample of the 91 children housed in Foundling Home, Pinneau found himself at a loss and therefore had recourse to the following conclusion: "It would seem obvious that in the original study the Foundling Home infants were tested only once, and that the original graphs were based on cross-sectional data." This statement is in contradiction to the facts presented in my studies on "Hospitalism." Pinneau makes this the backbone of his arguments. He variously refers to "the brief" period of time, to "the extremely short time" spent in the observation of Foundling Home, to "the purely cross-sectional nature of the study," and thus he progressively raises his own bid to the point where he has convinced himself that my observation of the children in Foundling Home was limited to "less than a week" and that these children were tested only once.

The *facts* are as follows: The children in Foundling Home were observed by me for over three months, daily, during four to six hours per day. During this time they were tested, filmed, and observed by myself and my assistants. In the course of the subsequent two years they were observed and photographed at four-month intervals by an assistant trained by myself (**17**, pp. 114 f., **23**, 270 f).

The technique used in the above two examples of Pinneau's argumentation is applied by him with many variations, modifications, and elaborations. A further example is provided when he indignantly states that the 91 children in Foundling Home were not followed for two years, *because in the course of these two years* 34 had died. He appears not to have grasped a simple medical fact: if the initial sample of a population of 91 is followed for two years, it is the mortality statistics and the condition of the survivors in which the physician is interested. The death of 34 of the original population will provide the physician with the most alarming proof that the conditions for survival are precarious indeed.

The physician will regret that another 36 of the original sample could not be followed because they were adopted or transferred; but he knows that this is one of the shortcomings of longitudinal studies. And the absence of data about these 36 can under no circumstances invalidate the horrifying finding that more than one-third of the original population died within two years. He will realize that there is every probability, even in the best of circumstances, that information about the 36 who successively disappeared from the original sample could only have added to the percentage of the mortality. To state under these circumthe purposes of a study on the first year of life longitudinal has to be defined as comprising a period sufficient to detect significant developmental changes in the subject. In the first year of life such a study will require at least two and preferably three months.

I will begin with Pinneau's treatment of the cases in the original Nursery group. He has confused the figures of the original sample of 69 children observed in Nursery for the purposes of the study on "Hospitalism" (**16**, **17**) with the sample of 123 children observed in the same institution for the purposes of "Anaclitic Depression" (**18**). He posits an arbitrary date for the beginning of observations in Nursery and thus succeeds in creating the impression that only one-third or at best one-half of the children in question in either sample were observed consistently for a full year. He concludes from this that "the graphs based on the sample of children must to a large extent represent different children at each age," a wording which creates in the reader's mind the impression that most of the children were seen only once or twice in the course of their first year.

I can dispel the confusion created by Pinneau. The 69 children who form the sample of the article on "Hospitalism" (16) consist of:

39 children seen for 12 consecutive months or more from birth (within two weeks of delivery);
15 children seen for 11 months or more, beginning within the first month of life;
9 children seen for 11 months, beginning within the first month of life;
6 children seen for 10 months, beginning within the second month of life.

Let me add for those not familiar with infant observation that no test of any kind has yet been devised to provide really useful psychological information on the infant in its first two months of life, during which behavior is more in the nature of reflex than response. Therefore we did not consider it particularly important that a minority of our sample was not seen by us in the first month and a few of them not in the second month. It might also be mentioned that when Pinneau discusses the usefulness of the Viennese test, most of the discrepancies which he makes so much of are caused just by the unreliability of all and any test scores in the first two months of life. Each of the 69 children was seen at least once a week during the whole period of his stay by myself and/or my associates. It seems to me that the 30 days by which some of the children fall short of the full year's observation hardly can detract from the longitudinal character of the study. When possible each child was tested once a month. When this was not feasible, bimonthly intervals were observed.

After having put in question the original sample of 69 children dealt with in "Hospitalism" (**16**), Pinneau casts doubts upon the procedures used with subsequent children in the Nursery. The fact is that the populations that were discussed in the studies in my subsequent articles ("Anaclitic Depression" (**18**), "Autoerotism" (**21**), "Psychogenic Diseases" (**23**)) were routinely observed beginning with the first day when the child arrived (that is, within two weeks of birth) at Nursery. A film record was made of its first responses, and the observations continued until the child left the institution, that is, after the completion of the first year.

Pinneau's argumentation on the quantitative aspects is therefore invalid when all the facts and not merely a selected portion of them are examined.

We will now proceed to examine other of Pinneau's treatments of data "to obtain the flavor" (as he puts it) of his criticism.

Perhaps this is the place to mention that a good many of Pinneau's interpretations were made possible by the special circumstances surrounding my research. The physician has a professional obligation of discretion to his patients. Furthermore, penal institutions are forbidden by law to give any clue to the identity of their inmates and the same goes for Foundling Home. I made every effort to keep identity, time, and place protected from uncalled-for inquisitiveness, to the point of using, here and there, misleading clues, such as the term "Western Hemisphere." For the reader's information that term was used by me in a cultural sense, meaning the Western world, including Europe.

When therefore I avoided publishing any specific date on the sample of the 91 children housed in Foundling Home, Pinneau found himself at a loss and therefore had recourse to the following conclusion: "It would seem obvious that in the original study the Foundling Home infants were tested only once, and that the original graphs were based on cross-sectional data." This statement is in contradiction to the facts presented in my studies on "Hospitalism." Pinneau makes this the backbone of his arguments. He variously refers to "the brief" period of time, to "the extremely short time" spent in the observation of Foundling Home, to "the purely cross-sectional nature of the study," and thus he progressively raises his own bid to the point where he has convinced himself that my observation of the children in Foundling Home was limited to "less than a week" and that these children were tested only once.

The *facts* are as follows: The children in Foundling Home were observed by me for over three months, daily, during four to six hours per day. During this time they were tested, filmed, and observed by myself and my assistants. In the course of the subsequent two years they were observed and photographed at four-month intervals by an assistant trained by myself (17, pp. 114 f., 23, 270 f).

The technique used in the above two examples of Pinneau's argumentation is applied by him with many variations, modifications, and elaborations. A further example is provided when he indignantly states that the 91 children in Foundling Home were not followed for two years, *because in the course of these two years* 34 had died. He appears not to have grasped a simple medical fact: if the initial sample of a population of 91 is followed for two years, it is the mortality statistics and the condition of the survivors in which the physician is interested. The death of 34 of the original population will provide the physician with the most alarming proof that the conditions for survival are precarious indeed.

The physician will regret that another 36 of the original sample could not be followed because they were adopted or transferred; but he knows that this is one of the shortcomings of longitudinal studies. And the absence of data about these 36 can under no circumstances invalidate the horrifying finding that more than one-third of the original population died within two years. He will realize that there is every probability, even in the best of circumstances, that information about the 36 who successively disappeared from the original sample could only have added to the percentage of the mortality. To state under these circumstances that the 21 children who were seen from beginning to end is too small a portion of the original sample for the purpose of concluding that deprivation of maternal care has a harmful effect, appears, to put it mildly, naive.

But what shall we say when we find that Pinneau considers that the 21 survivals constitute "an inferior biological selection"! Does Pinneau really mean to convey that the 37 who died were biologically superior to the 21 who survived?

He raises many other objections which serve to becloud the issue. These appear to me less essential and I will deal briefly with a few of them only.

a. There is an implication that I have attributed the decline of 33 points in the developmental quotient to the consequences of weaning. He supports this allegation through quotations which, taken in their right context, mean just the opposite. On the contrary, I stated in several of my publications that the effect of weaning, particularly around the sixth month, usually results in a rise of the developmental quotient. On the other hand, throughout my publications I stressed that it is the deprivation of maternal contact and of reciprocity with the mother which has destructive consequences. Therefore weaning *as such* is not important; but when children are weaned, they concomitantly lose that contact with their mothers which is implicit in the feeding situation.

b. He questions the value of the Hetzer-Wolf test (**3**, **8**). But contrary to Pinneau's allegations it has been applied to many thousands of infants in many countries, is being used at present, and compares favorably indeed with all other tests (for the first year) known to me.[1] Regarding its predictive value, just as with any other test, a single administration is uninformative and only the trend of a number of successive applications of the test over a period of time is meaningful.

c. He questions the composition of my staff: It consisted of myself, Katherine M. Wolf (one of the two originators of the Hetzer-Wolf test), and of a number of assistants with Ph. D. qualifications trained by Wolf and myself in testing and observing infants.

d. He objects to the method of comparing different institutions with one another, a debatable point. But it is typical of Pinneau's procedure that he neglects to mention that his objection applies only to one of the five studies discussed by him ("Hospitalism" [**16**]) and that in the remaining studies, groups of infants present at the same time in the same institution, namely in Nursery, were compared with each other.

e. He objects that the question of congenital abnormalities was not touched upon by me and highlights this as one of the uncontrolled variables in the sample. But the nature of the institutions themselves implies that congenital abnormalities were excluded on admission, as the institutions in question were not equipped to deal with them.

f. He alleges that I have not employed normal groups of infants for purposes of comparison. That idea of Pinneau's is a negative hallucination. Within the publications examined by Pinneau, in "Autoerotism" (**21**) on page 103, there is a chart filling half the page, and showing the relationship between *normal* infants, infants in Nursery, and infants in Foundling Home—not to speak of the fact that my whole research is based exactly on this kind of comparison.

---

[1] Katherine M. Wolf informs me that she does not intend to answer Pinneau's attack on the Hetzer-Wolf test. She further calls the fact to my attention that the number of children to whom the Viennese test has been applied under controlled conditions can be found in Charlotte Bühler's book: "Kleinkinder Tests" (**3**) and in the publications which deal with the restandardization of this test in social and cultural groups, e.g., Hofstaetter (**9**, **10**), Maria Wolf (**26**, **27**), etc. My reference to thousands of children was based on the number of children tested in the course of the years of activity of the Viennese University Psychological Institute and elsewhere, without the results of these tests having been published.

This refutation of the objections raised by Pinneau could go on and on. I will refrain from going into further details, except to point out again his unfamiliarity with the subject, which becomes embarrassingly evident when he takes me to task because I mention marasmus as a cause of death only in ulterior publications and not in the first ones. He compounds this blunder in a note (footnote 24) where he points to a similar negligence on my part in regard to cachexia. He appears to believe that the two terms serve to designate distinct disease entities. They do not. Marasmus is a symptom, a progressive wasting away, especially, in infants, but also in senescents. The term cachexia also means progressive wasting away when used alone. As an adjective of the "more than a dozen conditions" which Pinneau mentions, it simply means the wasting away occurring in that particular disease. Pinneau takes the secondary effect for the disease.

The lack of clinical orientation is particularly evident whenever Pinneau tries to apply to clinical observation and to clinical methods his statistical standards. He introduces into the observation of the living being an atomistic concept and does not realize that the shortcomings of a test, with random fluctuations of 5 to 15 points are very small matters indeed, when in the follow-up the children in Foundling Home fall to an average level of 55 points below the "normal" average of 100, and when the essential point which I made is that these emotionally deprived children were dying in a truly horrifying percentage.

Pinneau's yardstick for test validation may be meaningful in evaluating the behavior of mice and rats; certainly his critical procedures will produce no harm there, for we are not responsible for the survival of laboratory animals and their species. The physician deals with human beings and their survival. I do not want to be dramatic or to harrow Pinneau's conscience. But this is not the first attempt by Pinneau to attack clinical fieldwork by purely deductive reasoning. He has tried to invalidate the pioneer work of Margaret Ribble (**14, 15**) and in the present case he also attacks Liselotte Fischer and Katherine M. Wolf. But his criticism of this line of research implies that we should continue to raise infants in such places as Foundling Home, at the risk of a mortality of 37½ per cent in two years, and wait for a change until we have ironed out the minor fluctuations in the various tests applied. The results of the work of pioneers like Chapin (**4**), Lowrey (**11**), Goldfarb (**6, 7**) and Bakwin (**1**) has fundamentally changed our approach to the raising of infants in the last 40 years. The results achieved by the later workers, Anna Freud (**5**), John Bowlby (**2**), Margaret Ribble (**14, 15**) and I believe I can include myself, has been by this time applied all over the world in the practice of many hundreds of hospitals and institutions, with a concomitantly demonstrable saving of innumerable human lives. The exertions of Pinneau will not stop the progress in the care and the understanding of infants achieved in the last 40 years. When faced with this kind of criticism, we can say with Henry Poincaré: "On ne discute pas avec un physicien, on répète ses expériences!" The quality of Pinneau's criticism has been aptly characterized in Joseph Stone's presidential address to the N.Y. State Psychological Association. There, in a few masterful paragraphs of comment on Pinneau's criticism of Margaret Ribble's work, Stone includes the remark: "I commend you his article [Pinneau's] as a kind of hydrogen bomb perfection of destructive criticism; not a paragraph is left standing for miles around" (**24**).

Nobody realized better than I the shortcomings of my experimental design, and those of my publications. They are largely due to the exiguity of the means at my disposal, to the space restriction imposed on such publications in psychiatric journals, and, last but not least, to the fact that I was exploring uncharted territory in which progress can only be halting. But Pinneau, after all, could have asked me to explain apparent inconsistencies and to fill out gaps, if it was information and the advancement of science he was after. He chose not to do this.

Criticizing, questioning, discussing empirical and nonempirical studies in science increases our insight, broadens our knowledge and is highly welcome when based on factual data. A critical discussion, however, built on inference and implication, and which has recourse to invention, serves no useful purpose. On the contrary; it beclouds the issue, it confuses and, moreover, holds up unnecessarily the progress of science.

REFERENCES

1. Bakwin, H. Loneliness in infants. *Amer. J. Dis. Child*, 1942, **63**, 30–40.
2. Bolwby, J. Maternal care and mental health. *World Health Organiz. Monogr. Series*, 1951, No. 2.
3. Buehler, Charlotte, & Hetzer, Hildegard. *Kleinkinder Tests.* Leipzig. Barth, 1932.
4. Chapin, H. D. A plea for accurate statistics in infants' institutions. *Arch. Pediatr.*, 1915, **32**, 724–726.
5. Freud, Anna, & Burlingham, D. *Infants without families.* New York: International Univer. Press, 1944.
6. Goldfarb, W. Effects of early institutional care on adolescent personality. *J. exp. Educ.*, 1943, **12**, 106–129.
7. Goldfarb, W. Effects of early institutional care on adolescent personality. Rorschach data. *Amer. J. Orthopsychiat.*, 1944, **14**, 441–447.
8. Hetzer, Hildegard, & Wolf, Kaethe. Baby tests. *Z. Psychol.*, 1928, **107**, 62–104.
9. Hofstaetter, P. R. Testuntersuchungen an japanischen Kindern und das Reifungsproblem. *Z. Kinderforsch.*, 1937, **46**, 71–112.
10. Hofstaetter, P. R. Was besagen Testergebnisse. Ein Beitrag zum Dimensionsproblem des Entwicklungstests. *Z. Kinderforsch.*, 1938, **47**, 92–96.
11. Lowrey, L. G. Personality distortion and early infant care. *Amer. J. Orthopsychiat.*, 1940, **10**, 576–585.
12. Pinneau, S. A critique on the articles by Margaret Ribble. *Child Develpm.*, 1950, **21**, 203–228.
13. Pinneau, S. The infantile disorders of hospitalism and anaclitic depression. *Psychol. Bull.*, 1955, **52**, 429–452.
14. Ribble, Margaret A. *The rights of infants: Early psychological needs and their satisfaction.* New York: Columbia Univer. Press, 1943.
15. Ribble, Margaret A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* II. New York: Ronald Press, 1944.
16. Spitz, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child*, **1**, 53–74. New York: International Univer. Press, 1945.
17. Spitz, R. A. Hospitalism: A follow-up report on the investigation described in Vol. I. *Psychoanal. Stud. Child*, **2**, 113–117. New York: International Univer. Press, 1946.
18. Spitz, R. A., & Wolf, Katherine M. Anaclitic depression. *Psychoanal. Stud. Child*, **2**, 313–342. New York: International Univer. Press, 1946.

19. Spitz, R. A. Environment vs. race. *Arch. Neurol. Psychol.*, **57**, 1947.
20. Spitz, R. A. The importance of mother-child relationship during the first year of life. *Ment. Health Today.* Wash. Soc. Mental Health, 1948, **7**, 7–13.
21. Spitz, R. A., and Wolf, Katherine M. Autoerotism. *Psychoanal. Stud. Child*, **3/4**, 85–120. New York: International Univer. Press, 1949.
22. Spitz, R. A. Three first steps in growing up. *Child Study*, 1950/51, **28**, 2–5.
23. Spitz, R. A. The psychogenic diseases in infancy. *Psychoanal. Stud. Child*, **6**, 255–275. New York: International Univer. Press, 1951.
24. Stone, J. A critique of studies of infant isolation. *Child Develpm.*, 1954, **25**, 9-20.
25. Wolf, Katherine M. Observation of individual tendencies in the first year of life. In M. J. E. Senn (Ed.), *Problems of infancy in childhood.* New York: Josiah Macy Jr. Foundation, 1953.
26. Wolf, Maria. Kleinkindertests. *Arch. ges. Psychol.*, 1935, **94**, 215–246.
27. Wolf, Maria. Kleinkindertests an Wohlstandskindern. *Z. Kinderforsch.*, 1935, **44**, 191–193.

REPLY TO DR. SPITZ[1]

SAMUEL R. PINNEAU
University of California

In replying to my evaluation of his research on infantile disorders, Spitz presents a number of statements which may be examined in relation to specific sections of the original review.

The Research Reports Reviewed

Spitz considers that in dealing with only five articles, I have used an incomplete sample of "three or four dozen scientific and professional articles" and "about a dozen motion pictures."

A search of the *Psychological Abstracts*, the *Quarterly Cumulative Index Medicus*, and the *Biological Abstracts* disclosed from 1935 to the present, 23 articles by Spitz and four films which report aspects of his research on infants.

My choice of articles was dictated by four considerations: Of the articles I read, these gave the most detailed consideration of the institutions, the infants, and their background. All were available in the same Annual (**13**). In American psychology, these are the most widely cited among his publications. In his references, he himself refers to them more frequently than to other reports.

The Studies

In his reply, Spitz has failed to give essential information about the nature of his studies. The Nursery study was apparently initiated in 1942 and that of Foundling Home in 1944 (**4**, p. 432), at which times his residence is listed as New York (**12**); data are still withheld concerning the social and geographic areas served by these institutions. The reply suggests additional questions concerning the comparability of the samples, inasmuch as Nursery is apparently in a penal institution in New York (**4**, p. 433), while the location of Foundling Home is no longer limited to the "Western Hemisphere," but rather to the "Western world." While a physician's responsibility to his patients or subjects is not to be denied, it is difficult to believe that in a matter of this sort professional ethics or legal restrictions could be violated by identifying the institutions concerned, or at any rate by giving de-

---

[1] The writer wishes to acknowledge the criticisms and suggestions of Harold E. Jones in the writing of this reply.

tails as to national, educational, and socioeconomic samplings involved.

THE SAMPLES

Spitz states in his first article (**5**, p. 56) that 130 infants constituted the *total* population of both institutions, of which 69 were in Nursery and 61 in Foundling Home. Elsewhere in the article the number in Foundling Home is stated as 88 (**5**, p. 59) and in a different article, 91 (**9**, p. 271). In the reply he refers to 91 infants in this institution, and states that "If the initial sample of a population of 91 is followed for two years, it is the mortality statistics and the condition of the survivors in which the physician is interested." Very true; but the fact remains that the number of cases in this institution actually followed for two years was 21 and not 91, unless all those that have died are automatically registered as members of subsequent follow-ups.

THE NATURE OF THE OBSERVATIONS

In his reply, Spitz states that his original study of the Foundling Home infants required three months. This clarification is important since his earlier statements, on which I based my estimate of the length of time required for the original study (**4**, p. 435), indicate repeated tests on these children over a two-year period (cf. **4**, p. 438). (The graph for this group is reproduced in Fig. 1 (**4**, p. 440)). While it must be granted that three months of observation can be considered a sufficiently long period of time to qualify the study as a longitudinal one for *some* purposes, it is quite obvious that it cannot be used to show the development of a constant number of Foundling Home subjects during their first year of life. Spitz still does not specify the number of children observed at each age, nor the number of times the individual children were tested. This may seem to be a trivial matter, but lack of information in such areas and repeated discrepancies, such as noted above, make it difficult to evaluate both the reports of the observations and the interpretations which are offered.

With regard to the Nursery infants, we are informed that the total sample in the institution was included (**5**, p. 56), that observations began in the first two months, and that each was observed for a period of 12 to 18 months (**7**, p. 313). Thus, at the time the study began on the original sample of 69 cases, there could have been no children in the institution beyond two months of age. But we know that Nursery had been in existence for at least ten years (**6**), that it admitted children at frequent intervals (**4**, p. 433), and kept them through their first year (**5**, **6**). A check on the earlier records of the institution might explain this apparent discrepancy, if the institution could be identified.

THE RESULTS

Clinical findings. As previously noted in detail, Spitz's clinical findings appear to me to contain numerous contradictions, (see, for example, **4**, p. 434, 437, 445, 446). In addition, the statements concerning the care of the children in the two institutions are contradictory (cf. **4**, p. 437–438), there are known differences between the two groups that could at least in part account for the results obtained (cf. **4**, p. 433–435, 447), and there is lack of information on a number of other variables which might account for the results (cf. **4**, p. 437, 447).

He does not deal with any of these specific points in his reply except that of congenital abnormalities. With regard to these he states that the nature of the institutions implies that they were excluded on admission; however, this statement is difficult to reconcile with his description of the Foundling Home infants' medical care (cf. **4**, p. 437–438; **5**, p. 62) and his statement that in advanced extreme cases the picture of these children varied from stuporous deteriorated catatonia to agitated idiocy (**7**, p. 331).

In his articles Spitz lists marasmus and cachexia among the causes of death of the Foundling Home children; however, in the reply he also maintains that these are just symptoms. The question arises as to whether these symptoms, as he now refers to them, were the cause of the death or if the cause was a "disease entity" of which marasmus was the symptom. Measles was listed as the cause of 65 per cent of the deaths; the particular disorders leading to death in the other 35 per cent were unspecified. It should be noted that the distinction between "symptoms" and "disease entities" is not always a clear one, and that apparently not all investigators in the medical sciences would agree with Spitz regarding maramus (**1**, **2**, **11**; cf. **3**, p. 221–222 for a summary of these papers).

Statistical findings. In his studies Spitz has used the Hetzer-Wolf test. An evaluation of the test in terms of published research indicates that it was inadequately standardized and that a given DQ does not have the same relative significance at different ages within the first year. The available evidence does not support Spitz's contention in his reply that the predictive value of this test is increased by having repeated measurements on the same children during their first year. While it is undoubtedly true that the test has been and is used extensively in many countries, frequency of use cannot, in and of itself, increase the validity of the test.

According to Spitz, the most prominent feature of hospitalism, manifested by the Foundling Home infants, was severe developmental retardation as determined by the Hetzer-Wolf test (**5**). In his reply he continues to maintain that the drop in DQ score is to be accounted for by the separation of the child from his mother. The average DQ of the children fell from approximately 131 to 72 in the first year; however, the drop from approximately 131 to 76 took place before the prevalent age at which the children were separated from their mothers (cf. Fig. 1 [**4**, p. 440]). In his first article (**5**) he states that by the end of the second year the DQ sinks to 45, " . . . an average level of 55 points below the "normal" average of 100 . . . " (**10**). It is to be noted, however, that the individual children on whom this statement was based were not tested over a two-year period, but only over a three-month period, and the DQ decline as reported must therefore involve different samples or different combinations of cases at different ages, the amount of repeated testing and of sample overlap not being specified.

The total Nursery sample of 196 cases was divided into three subsamples in my original article: the original group, the anaclitic depression group, and a residual sample. It was suggested that Spitz chose those subjects showing a marked drop in score for the anaclitic depression sample. Until he has specified that this is incorrect, and why, it can only be assumed that the results in this study can be accounted for in this manner.

SPITZ'S CONCLUSIONS

Spitz lists three conclusions in his

reply which he feels that I have challenged. I have, however, not challenged these as propositions, but have merely raised questions as to whether they are in fact supported by the evidence which he presents. He says that my evaluation of his research " . . . implies that we should continue to raise infants in such places as Foundling Home . . . " (10). This is incorrect. The purpose of a critical evaluation of a research report is to determine if the data reported warrant the conclusions drawn by the author. I do not believe that Spitz's data do.

CONCLUSION

Dr. Spitz's articles are almost the only research reports in the first six volumes of *The Psychoanalytic Study of the Child* (**13**) which attempt a statistical analysis of the data. He is to be commended for this; certainly in this area of clinical hunches and hypotheses we have special need for more careful scientific investigations and checking of hypotheses, and the use of statistical materials may be a step in this direction. It may well be that the burden of blame for the uncritical acceptance of his works does not rest with Spitz, who has published his results as he sees them, but rather with those who have acclaimed his work, and whose research training should enable them to make a critical evaluation of such research reports.

REFERENCES

1. HISLOP, W. A. Treatment of marasmus by injection of an extract of adrenal cortex. *Lancet*, 1938, **235**, 308–310.
2. HOLT, L. E. *The diseases of infancy and childhood.* New York: Appleton, 1918.
3. PINNEAU, S. R. A critique on the articles by Margaret Ribble. *Child Develpm.*, 1950, **21**, 203–228.
4. PINNEAU, S. R. The infantile disorders of hospitalism and anaclitic depression. *Psychol. Bull.*, 1955, **52**, 429–452.
5. SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child*, **1**, 53–74. New York: International Univer. Press, 1945.
6. SPITZ, R. A. Hospitalism: A follow-up report. *Psychoanal. Stud. Child*, **2**, 113–117. New York: International Univer. Press, 1946.
7. SPITZ, R. A. Anaclitic depression. *Psychoanal. Stud. Child*, **2**, 313–342. New York: International Univer. Press, 1946.
8. SPITZ, R. A. and WOLF, KATHERINE M. Autoerotism. Some empirical findings and hypotheses on three of its manifestations in the first year of life. *Psychoanal. Stud. Child*, **3/4**, 85–120. New York: International Univer. Press, 1949.
9. SPITZ, R. A. The psychogenic diseases in infancy: An attempt at their etiologic classification. *Psychoanal. Stud. Child*, **6**, 255–275. New York: International Univer. Press, 1951.
10. SPITZ, R. A. Reply to Dr. Pinneau. *Psychol. Bull.*, 1955, **52**, 453–459.
11. ANONYMOUS. Infantile marasmus. Current comment. *J. Amer. Med. Ass.*, 1930, **95**, 1, 538–539.
12. *American Medical Directory.* (17th and 18th Ed.) Chicago: American Medical Association, 1942, 1950.
13. *The psychoanalytic study of the child*, 1–4. New York: International Univer. Press, 1945–1951.

Good, Carter V., & Scates, Douglas E. *Methods of research.* New York, Appleton-Century-Crofts, 1954. Pp. xx+920. $6.00.

This is a lengthy volume addressed to "field workers, graduate students and members of the senior division of the undergraduate college who would evaluate the quality of conclusions, either as producers or consumers of research." It carries the subtitle: "Educational, Psychological, Sociological," an indication of the authors' aim to furnish a treatment of certain research methods common to education and the social sciences. Although the main effort is addressed to students in education, the subtitle is justified in a limited degree, both by the kinds of methods which are discussed and by the use of illustrative materials drawn from psychology and sociology.

Two preliminary chapters, one entitled "Research as a Way of Progress," and the other, "Formulation and Development of the Problem; Research Programs and Needs," introduce the main subjects. These are: literature and library techniques; the historical method; the descriptive method (considered from the points of view of analysis, classification and normative research); the experimental method; methods used in case and clinical studies, and in genetic and developmental studies; and the reporting and implementation of research.

From the point of view of the needs of students of psychology, whether undergraduate majors, graduate students, or established scholars in the field, the values in the book are somewhat uneven, and in a certain sense collateral. The first chapter on "Research as a Way of Progress," for example, is uncritically inspirational, being based on the thesis that "such strength as will be demanded for survival and for the good life can come only through research; not primarily research in the physical sciences—granted that such research must be continued on a large scale and probably made more fundamental—but research on all of the many facets of living." Now this is an unobjectionable sentence, if an uncritical view is taken of the meaning of the word *research.* However, if by research is meant methodology, particularly methodology as it is defined by the content of this book, the sentence is debatable since it leaves unmentioned the influence of the theorist and, to pass to another universe of discourse, those forces in the development of knowledge which seem to function like the forces closing a gestalt—"discoveries" have been made, it has often been noted, because the time was ripe. In such cases, the men were less important than the ideas with which they were working.

The main treatment suffers from a certain dilemma in which the authors are caught. They do not, they indicate, want to write a textbook on techniques—statistical, psychometric, or sociometric—and neither have they defined their audience as one which can be expected to be competent in the use of such techniques. On the other hand, they are obliged for various reasons to come to terms with all the ways in which problems of research are problems in statistics, in experimental design, or in validation—to give but a few illustrations. The results are not always happy. Their aim is accomplished in too many instances, of importance to the training of psychologists, by mere mention of the names of techniques and methods—sometimes in lists which convey nothing but the name—by referring the student to other sources, or by passing judgment upon the applicability of techniques in an authoritative way. (No serious student should be left depending upon a professor's authority when it comes to such a matter as the applicability of a technique.) Too seldom is an account offered of their use which illuminates the problems for which they are suitable, or gives detailed guidance in their use, or alerts the student to the special difficulties which must be overcome in their application, or most important of all, leaves the student with a firm grasp of the internal logic of the method which fits him to be his own authority.

The work is strongest where it is sensitive to needs which are either shared by all three kinds of students to whom it is addressed (in education, psychology, and sociology) or to the needs of the students of education alone. Thus it contains a useful treatment of the methods of the historian—about which psychologists ought generally to know more—with practical guides to historical sources, to note-taking methods, to the very important techniques of criticism, both external and internal which are the special virtue of the professional historian, and to the problems of historiography. It contains a useful, but brief, account of questionnaire and interview techniques, and of the problems encountered in the conduct of the kind of educational survey with which professional educationists are familiar. More superficial, but still useful, are the accounts of content analysis, small-group study, and the conduct of research in association with casework.

The book has a number of collateral values which make it a useful reference work. First are the extensive bibliographies. These are organized as footnotes (of which there are nearly 1,600, many of which contain dozens of individual citations), and as Selected References (of which there are nearly 2,100). It is quite possible that the book contains more than 6,000 citations in all. Further, the various bibliographies abound with references as recent as 1954. Second are the many side references, principally of an historical turn, which are a tribute to the wide-range interest of the authors and which enrich the reader—we learn of Hull's practice of keeping a notebook in his graduate days at Wisconsin, that at the end of his career the notebooks totaled twenty-seven volumes, and that some of his stimulus to systematic thinking came from these notebooks, and similarly of Ranke, Mommsen, Huxley, Boring, Tolman, Hall, Bingham, Niebuhr, Sumner, Giddings, Ehrlich, Gibbon, and Kettering, who are merely a few among many others.

Despite these values, however, the work carries implications which will trouble many readers. It implies the primacy of data collection and treatment over the process of reflection from whence come the theories and hypotheses which direct the choice of data to be collected. It carries implications that all kinds of data collection are equally respectable, from an intellectual point of view, as indeed they are where *only* procedural questions arise. But this latter assumption blurs important distinctions in the minds of students. The engineers know better than this when they distinguish explicitly between research and development. It implies that the student can be trained to "evaluate the quality of conclusions" on the basis of acquaintanceship and reference knowledge, i.e., knowledge of where to go in order to acquire the competency which will make one capable of asking the significant questions in the evaluation of the quality of conclusions. All of this seems to the reviewer to be likely to produce bystanders rather than participants in scientific work.

MALCOLM G. PRESTON
University of Pennsylvania

BRAND, HOWARD. (Ed.) *The study of personality: a book of readings.* New York: Wiley, 1954. Pp. xvi +581. $6.00.

The Study of Personality, as its subtitle accurately reflects, is a collection of readings, consisting mainly of papers that have previously appeared in journals. The editor has contributed four original chapters, including a general introduction and an introduction to each of the three sections into which the book is subdivided; he has also contributed a brief commentary preceding each of the reprinted papers.

The three sections of the book deal, respectively, with theory, methods, and problems. For inclusion in each of the sections the editor has selected material "from clinical, experimental, and social psychology, from anthropology, and from sociology." In his preface, he has indicted his awareness that such broad coverage has been achieved at the expense of omitting many important papers in the field of personality. However, as his avowed purpose was to "show the student how much variety there is in research activity touching on personality," the editor has not committed himself to any selection criteria in terms of which his choices can be criticized. One can only comment, in this connection, that he has selected an impressive collection of high-level treatments of personality, which have been written from many divergent points of view. The list of contributors reads like a condensed "Who's Who in the Social Sciences."

The editor's emphasis is mainly methodological, and even the section dealing with theory stresses principles of theory construction rather than attempts to formulate a single logically consistent theory of personality. This methodological emphasis, in the reviewer's opinion, definitely restricts the "student" group for which the book can serve a useful purpose; the group of potential readers is probably limited to graduate students and social scientists who have completed their formal training. Graduate students should be able to utilize it as a nucleus around which to organize their reading of significant research contributions to personality theory. Fully trained social scientists will likely appreciate its value in providing cues to the recall of significant contributions that have undergone a process of partial forgetting.

Although the editor does not explicitly state that his book is expected to be of use in the teaching of undergraduates, he does not disclaim its usefulness for this purpose. On this point, the reviewer is inclined to believe that, while the book may become a rich source of lecture material, it cannot be understood adequately by the average undergraduate reader because of its level of sophistication, its level of difficulty, and its specialized emphasis.

The clinical psychologist, and the instructor with clinical inclinations, will likely find this collection of readings somewhat disappointing. There is little in it to satisfy his ideographic propensities, since the attitude of the editor appears to have been that of assuming that the scientific method in psychology requires a nomothetic approach. That ideographic approaches can be just as "scientific" as any of the nomothetic approaches has been amply demonstrated in the psychological literature. The reviewer, at any rate, is in full agreement with the clinical psychologists and psychiatrists who insist that the investigation of a single personality must be a process of entertaining, testing, and confirming, modifying, or rejecting hypotheses which develop out of the data as they are being collected, whether the data consist of responses to diagnostic tests or responses occurring during psychotherapeutic interactions.

In summary, this collection of readings should fill the need for supplementary reading in a graduate level course dealing with nomothetic approaches to the investigation of personality, or as a supplement to a course in the theory of personality when either the instructor or a good textbook contributes a single consistent theoretical point of view. Psychologists and other social scientists will find it useful as a review of significant and more or less familiar material that should not be allowed to become lost in seldom consulted back issues of the journals.

BERT R. SAPPENFIELD
Utah State Hospital

KERMAN, EDWARD F. *What is electroshock therapy?* New York: Exposition Press, 1954. Pp. 152. $3.50.

Although biased in favor of electroshock, this book should serve a useful purpose in supplying nontechnical authoritative answers to questions usually asked about this popular form of treatment by patients and their relatives.

JAMES D. PAGE
Temple University

RYAN, THOMAS, & SMITH, PATRICIA. *Principles of industrial psychology.* New York: Ronald, 1954. Pp. xiv+534. $5.50.

According to Ryan and Smith, "the book is designed as an introductory survey of the entire field of industrial psychology." They are "concerned with formulating for the reader the principles of industrial psychology."

Since the authors omit many topics, this book cannot be considered a survey of the entire field. In addition, the reader may have some difficulty in identifying the principles from the textual material and so it appears as if this is more a title of a book rather than a clearly stated and outstanding series of principles that lead to a systematic presentation of industrial psychology.

The book presents in substantial fashion those aspects of industrial psychology as the authors perceive it to be. While one may disagree with their point of view, it is clear that the authors are very familiar with the subject matter.

The style of presentation is characterized by critically evaluating research studies reported in the literature and emphasizing the necessary statistical concepts and techniques related to selection of employees. The heavy statistical involvement may make this book a little too difficult for the typical undergraduate student who is not a psychology or statistics major. Admittedly, statistics is an important tool of the industrial psychologist, but whether statistics and industrial psychology can be learned simultaneously from the same text is questionable. One may also wonder whether the student will become confused as a result of the authors' criticisms of other researchers' work. Many references seem to be included to illustrate shortcomings of research in the field. This is contrary to the usual practice in basic texts of referring only to the best work done in the field and emphasizing the positive aspects of the findings. A completely interested (possibly graduate) student will benefit considerably from such a presentation. The ordinary ones may become annoyed.

Part I, "Selection and Placement," is a solid and conservative presentation of tests and other selection procedures with a considerable emphasis on statistics. Slightly more than 50 per cent of the entire book is devoted to selection and placement. Part II is devoted to "Factors in Efficiency." The first sentence states that "All of this book is concerned with the ways in which efficiency of the performance of workers can be improved." Ryan and Smith apparently view this as the main task of industrial psychology. In addition to considering the concept of efficiency, such environmental factors as lighting, ventilation, noise and their effect on efficiency as well as a critical chapter on time and motion study are included.

There are three chapters on industrial motivation in this book. The authors reject what they term the behavioristic and the Freudian biases and propose a general theory on the initiation and control of activity. They state "that man behaves with respect to the future and past as they anticipate or remember them." According to Ryan and Smith, "the first stage in explaining behavior is a problem of perception, remembering, imagining, understanding and thinking."

From this take-off point, a series of thirteen hypotheses are postulated in the second chapter on motivation. Two examples are: Hypothesis I—Every individual wishes to believe or perceive himself as valued by the social group as an improved and useful member; Hypothesis IX—Developing skill in any activity tends to enhance the goal character of the activity.

The third chapter on motivation covers such topics as job satisfaction, attitudes, morale. Brief references are made to the Hawthorne studies, the work at the University of Michigan, and experiments done by some of our English colleagues. In addition, boredom, monotony and restriction of output are mentioned. Additional chapters on facilitating learning and accident control conclude the authors' coverage.

Ryan and Smith view the contributions of industrial psychology as (*a*) placement, (*b*) evaluating efficiency, (*c*) motivation, (*d*) training methods, and (*e*) accident proneness. If one is content with viewing industrial psychology as the above-mentioned areas and if one prefers to have slightly more than 50 per cent of the text on placement, then this text will be very satisfactory. There is a sound and critical presentation of the material covered. The major novelty of the text is the presentation of a theory of motivation. Whether it "catches" remains to be seen. From a reviewer's point of view it seems to be a description of the ways people might behave rather than a systematic presentation of causality. It does have shades of Woodworth, Gestalt, and Tolman and quite a point is made in rejecting Watson and Freud.

Principles of Industrial Psychology is an interesting book for a sophisticated audience. It may be misunderstood by typical undergraduates and it may not be too appealing to the man in industry who wishes to apply some principles.

MILTON L. BLUM
College of the City of New York

OESER, O. A., & HAMMOND, S. B. (Eds.) *Social structure and personality in a city.* New York: Macmillan, 1954. Pp. xxii+344. $4.50.

OESER, O. A., & EMERY, F. E. *Social structure and personality in a rural community.* New York: Macmillan, 1954. Pp. xiii+279. $3.75.

This two-volume overview of the sociological and psychological structure of the Australian Commonwealth, its people and its institutions, is admirably unique in many respects. The reported research was initiated by the Faculty of Psychology at the University of Melbourne, and aided by local and Unesco grants for the study of social tensions. It is important to note, however, that the work was carried on not only to render contributions to scientific knowledge, but to serve also as instructional devices for University classes in Collective Behavior (Social Psychology). Quite apart from the value of the material presented, it stands as a tribute to a style of cooperative student participation in professional research and reporting that would be unusual anywhere. Considering the reported personal autonomy of Australians, it is the more remarkable there.

It is not uncommon for single research programs to be oriented to the perceptual framework of the respondents used in the study. Indeed, this is quite characteristic of field studies in sociology and cultural anthropology. It is uncommon, however, to find this frame of reference saturating the psychological exploration of a total national culture as is done in this project. The reader is, for example, spared a dreary breakdown of class and caste distinctions as deduced by the investigators. Instead he is exposed to the self-revelations of the population with respect to their perceived roles and their role-tied assessment of institutions and social strata. Early in the first volume the reader is prepared for this more functional method of interpretation by a so-called "life phase diagram." The diagram schematically emphasizes individual growth, not as a function of chronological age, but as the phase-like passage through institutional groups. The adjustment of the individual to these social settings, his interpretation of them, and his evaluations of their interrelationships with him form the nucleus of the research methodology. The advantage gained by this approach is quickly apparent in its highly objective, yet never coldly detached, portrayal of Australian "culture." One looks in vain for symptoms of creeping national pride, or its reverse—the sterility of "guinea-pig" reporting. Neither is grossly apparent, and the native Australian would quickly identify with the detailed portrait of himself drawn by the authors.

There are many surprises for the American reader who "knows" Australians as a result of war-time experiences, single contacts with visitors, or Sunday supplement stereotypes. He will discover the Australian to be quite different from himself in expressions of nationalism, in "democratic" ideology, in cooperativeness, in neighborliness, and in a host of specific social behaviors. This reviewer would confidently recommend Volume I as a psychological Baedecker for the American traveler in urban Australia. Volume II may be no less authoritative as a rural guide, but it appears to be somewhat more academic and "reasoned" in its treatment of the material.

Without prejudice to the other contributors, special mention should be made of the extraordinarily fine development of the material by S. B. Hammond, and the chapters in which

P. G. Herbst details his conceptualization of family structure. It is Hammond who gives the initial orientation to the inquiry, and who acquaints the reader with the necessary background of history, existent social institutions, and the general way of life. At best this is a tricky job—and at worst is apt to be superficial. Hammond has avoided the attendant dangers and, in a lucid unfolding of the theme, paves the way with both general observations and specific reference to research data. In addition, his interpretive summaries seem to jell what might otherwise be scattered and sparse information.

Herbst's contribution is of another sort. Using a topological description of family structure, he has provided —at least within the framework of the Australian culture—a conceptual tool for characterizing the family both as a field of activity regions (through which family members pass developmentally) and as a locus of interaction patterns for family members. Family tension is dynamically linked with family structure. This is no esoteric abstraction, as handled by Herbst, but an experimentally implemented system which holds promise of a definitive comparison of cultures.

Despite these many attributes, the volumes do have certain shortcomings which detract somewhat from their readability. There is apparent in them a degree of disorganization and asymmetry that is at times disconcerting. One gets the impression that although the general editor did yeoman duty in connecting successive reports, the theme development and writing style of individual contributors did not always lend themselves to smooth internal blending. Volume II, dealing with rural community patterns, is better organized, but fails to parallel as well as it might the techniques and types of interpretation found in the volume on urban structure.

What appears to be another defect —instances of limited samplings of a total population of over eight million people—is actually not as damaging as it first appears. The studies of Hammond and Herbst, for example, show an extremely high consistency of behavior patterns apparent in family relations. Participation patterns of husband and wife are predictable with 92 and 95 per cent accuracy, respectively. In a culture where behavior is sufficiently homogeneous —as appears to be the case here—the sample size becomes less crucial.

Finally, there is an inconsistent policy of statistical reporting. Throughout Volume II there are adequate indices of significance appended to most tables. In Volume I, however, one must look long and hard for any such criteria.

In no sense do these limitations seriously lessen the worth of the books for a rather wide assortment of potential readers. For the psychologist, sociologist, anthropologist, and political scientist, they should constitute an invaluable reference book, both for content and methodology, in the delineation of national mores and their psychological origins.

DAN L. ADLER
San Francisco State College

BURT, CYRIL. *The causes and treatment of backwardness.* (Rev. Ed.) New York: Philosophical Library, 1953. Pp. 128. $3.75.

This book is a revision and expansion of earlier lectures and writings of Dr. Burt. "The subnormal child has already formed the subject of numerous inquiries and researches; and it seemed to me that the most useful thing that a psychologist could do would be to survey what is already known and what has already been accomplished, and then to summarize the main conclusions in a form available for the ordinary teacher." Although Dr. Burt's survey is fairly incomplete, it could be read with some profit by the grade school teacher.

SEYMOUR B. SARASON
Yale University

KUGELMASS, I. NEWTON. *The management of mental deficiency in children.* New York: Grune & Stratton, 1954. Pp. xii+312. $6.75.

This is an attractive, well-manufactured book for the guidance of professionally trained clinicians who meet handicapped children in their practice. The greater part of the book is devoted to a description of the syndromes. An introductory section on diagnosis and a final chapter on management constitute the chief difference between this book and a standard text on mental deficiency.

I am of two conflicting opinions. First, one cannot help being irritated by an extremely careless job of editing. One can't escape the feeling that such sloppiness might carry over into the crucial material of the book. There are instances of its having done so. On the other hand, if one can ignore these faults and consider the book as a whole, and in terms of its objectives, one must say that it contains a great deal of valuable information.

The classification scheme groups all disorders under four major varieties: developmental, metabolic, neuromotor, and psychological. Each of these varieties includes several specific syndromes which are described in some detail in terms of etiology, symptoms, and, in some cases, medication or treatment. While this etiologic grouping would seem to be satisfactory, the reader may be confused to find deficiencies due to cerebral injury, sensory damages and nutritional disorders classified among the psychological varieties.

The author shows much greater knowledge and understanding of medical, physiologic, and pediatric matters than in psychiatric or psychological areas. The book is full of flat statements: some are misleading, some are simply wrong, and some are silly. It is hard to understand this kind of writing and editing. It will certainly alienate the professional reader who knows something of the subject and it may mislead and confuse the uninformed who most need a book of this sort.

KARL F. HEISER
Louisville, Kentucky

EDWARDS, ALLEN L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954. Pp. xvii +542. $6.50.

This book is outstanding among the increasing number of texts designed to develop applied statistical competence in the perennially mathematics-free student of psychology, education, or sociology. Writing in conversational style, the author unfolds an extensive array of topics with maximum palatability and minimum sacrifice of modern statistical rationale.

The book is fairly large, containing 19 chapters followed by a 105-item bibliography, a list of 303 formulas cited in the text, an appendix of 15 tables, and answers to the excellent assortment of exercises presented in connection with each chapter. This all adds up to more than enough material for the first year of statistics.

As must be true of any introductory text, a great deal of traditional material is covered. A new book's claim to attention should be judged in terms of such criteria as clarity of exposition, self-containedness, selection of basic and specialized topics, provision of visual aids, examples, and exercises, relevance to problems of experimental design, and liaison with more advanced treatments. On all such criteria the book stands up exceptionally well. Conventional material on central tendency, variability, graphic devices, standard scores, normal curve, simple correlation and regression, other measures of association, chi square, and rudimentary analysis of variance is presented in rich detail. The only obvious omission is partial and multiple correlation and regression.

Contentwise the reviewer has only one question. This occurs in connection with Chapter 17 which deals with analysis of variance of a two-factor design. The first part of this chapter is devoted to analysis of a design involving the influence of three methods of instruction (lecture, discussion, project) upon three different kinds of achievement (a test of factual information, a test of understanding of general principles, and a test of ability to make applications). Although the tests are stated to be "comparable" it is didactically inappropriate to furnish the elementary student with an example where the dependent variable is made up of scores on three different tests. In addition the author describes the initial step in analyzing such a 3 by 3 design as testing the differences among the nine subgroups. In the example these differences are found to be significant and the analysis proceeds to the factorial form but the implication is strong that one should not go on to the factorial analysis unless the over-all differences are found to be significant.

Over and above the usual material Edwards provides many attractive features not commonly found in the introductory text. Among these are sections on nonlinear curve fitting, the power function in tests of significance, one-tailed vs. two-tailed tests of significance, combination of tests of significance, and Tukey's procedure for comparing individual means in conjunction with analysis of variance. Perhaps the most valuable special feature is the extensive presentation of nonparametric methods, a number of which are described in the same chapter with the analogous classical method while others are discussed in a final chapter on significance tests for ranked data.

LEONARD S. KOGAN
Institute of Welfare Research
Community Service Society of New York

LE BEAU, J. *Psycho-chirurgie et fonctions mentales.* Paris: Masson et Cie (Eds.), 1954.

This is an unusual, and in many ways, an exciting book. Written by one of France's leading brain surgeons, it deals with the general topic of psychosurgery from many angles. The author treats all the anatomical, physiological, clinical, and psychological principles involved, as well as different surgical techniques, particularly those related to selective ablations, medical complications, and postoperative treatment. Much space is given over to the results of psychosurgery in the treatment of neuroses, psychoses, mental disorders associated with epilepsy, the mental difficulties of children, and intractable pain. On all these points the author, drawing on his rich experience, has much of importance to say; his own material is well integrated with his surveys of the literature. The main importance of this work on the surgical side derives from the author's pioneering efforts in the direction of selective surgical ablation, in which he has followed in the footsteps of Clovis Vincent.

To the psychologist,.these achievements are of only peripheral interest. To him the main contribution of the book will center in those later chapters dealing with the psychological aftereffects of different types of operation. The importance of Le Beau's approach is twofold. In the first place, he does not rely, as so many other surgeons have done, on introspective reports and casual observation of the behavior of his patients. Nor does he give credence to results obtained on projective tests of low reliability and unknown validity, such as the Rorschach or the TAT. He argues firmly in favor of exact measurement in terms of objective tests measuring factorially ascertained dimensions of behavior. In this he follows the path indicated by A. Petrie in her book *Personality and the Frontal Lobes.* (Indeed, he makes specific reference to the help received from her in setting up his new psychological laboratory.) It is certainly a welcome change to find a surgeon knowledgeable enough to devote several pages to a discussion of the contributions of Thurstone, Cattell, and other factor analysts, and objective enough in his outlook to be willing to put his hypotheses to the experimental test in terms of objective personality measures. These hypotheses themselves are of very far-reaching importance and take us back to an earlier and more hopeful stage of psychological and physiological work, when expectations were high that mental functions could be located in certain parts of the cortex. By tracing out the detailed behavioral aftereffects of operations involving different parts of the cortex, and relating these changes to personality dimensions identified in terms of factor analysis, Le Beau has given us extremely interesting hypotheses, linking such factors as extraversion, neuroticism, etc., with definite Brodmann areas (p. 380).

It would be easy to criticize the picture given by him as naive, premature, and oversimplified. All this would undoubtedly be true. Nevertheless, such feelings should not keep the reader from studying the evidence in detail. It is surprising to note to what extent many apparently unrelated factors fall into place when seen in terms of the scheme presented by Le Beau. In any case, whatever the faults of the scheme, it possesses the outstanding advantage of being definite enough to permit of disproof; many deductions can be made from it which are experimentally testable. At an early stage of the development of a theory it would be unfair to ask for more than this: that it should unify known facts and predict unknown facts. It is to be hoped that this book will be widely read, and, equally important, that it be translated into English.

H. J. EYSENCK
University of London

VOL. 52, NO. 6 NOVEMBER, 1955

Psychological Bulletin

SOUND-PRECIPITATED CONVULSIONS: 1947 TO 1954[1]

WILLIAM BEVAN
Emory University

Sound-induced convulsions are still a scientific enigma. In the more than six years since Finger's (**27**) excellent review, nothing has been constructed that approximates a satisfactory explanatory framework. This does not mean that the phenomenon has lost its challenge. In the period covered by the present paper, approximately 129 audiogenic seizure titles were available to this reviewer. A check through Finger's bibliography revealed 106 titles, going back to the earliest reports of the syndrome. But there have been shifts in interest. The grinding of theoretical axes has more or less ceased and attention has turned toward cataloging—somewhat unsystematically—the variables that influence seizure susceptibility. Less of the work is being done by psychologists. Of the 106 papers cited by Finger, 92 were by psychologists, 14 by nonpsychologists; 90 appeared in psychological publications, 16 did not. In contrast, of the 129 items appearing in the past six years, 66 were authored by psychologists, 63 by nonpsychologists; 59 appeared in psychological publications, 70 were in nonpsychological journals. All of Finger's references were to American or Canadian journals; of those added after 1947, 19 have appeared in European publications. Recent work varies greatly in quality: some studies involve insufficient numbers of animals, inadequate design, and/or inept handling of behavioral data. Other experiments are exceedingly well done. Perhaps the most promising developments are concerned with the inheritance of susceptibility in mice, stimulus priming, and learned control of the seizure. The predominant theoretical bias is still physiological, but there is increasingly clear recognition of the overly simplified nature of earlier interpretations and of a need for a view flexible enough to accommodate the complex of data recently acquired.

The present review was planned as a supplement to Finger's. Thus it has been given approximately the same form.

PROCEDURES FOR TESTING AND FOR ANALYZING TEST BEHAVIOR

There has been little change in testing procedure since the early years of audiogenic seizure research. Experimenters most frequently use a small wash tub or other container of similar size as a test chamber. The most widely used sound-source is a door bell generating 90- to 100-db noise. Some workers, for a variety of special reasons, use pure-tone stimuli produced by a variable-frequency oscillator or compressor-driven Galton whistle. Few, if any, continue to use the air blast or jingling keys. The duration of the sound stimulus seldom exceeds two minutes, since the probability of a response after this time is small (**124**). Animals are usually allowed free movement within the confines of the test chamber, since complete restraint has generally been found to reduce sensitivity. Since too frequent testing has a similar effect, tests in a series are usually spaced at two- or three-day intervals. Some experimenters, however, have tested as frequently as twice daily (**18, 37**). A few workers (**37, 54**) include observations of the subjects' (*S*s') behavior for a short period prior to presentation of the auditory stimulus.

---

[1] The writer is indebted to J. L. Fuller of the Roscoe B. Jackson Memorial Laboratory, Bar Harbor, Maine, for reading critically the first draft of the present manuscript.

Despite criticism (**101**) that the stimulus lacks sufficiently precise definition, little concern has been directed toward further specifying the physical properties of the convulsion-provoking sound or clarifying the significance of the nonauditory characteristics of the test situation. Only fairly recently has a formal attempt at standardization of test conditions been made (**36, 124**). Because of their interest in audiogenic seizures as a technique of physiological and genetic assay, the division of Behavior Studies of the Jackson Memorial Laboratory at Bar Harbor, Maine, has undertaken to serve as a clearing house for information on seizures. This group recommends a two-minute test in the bell-wash tub apparatus with 30-day old mice, bred for sensitivity or immunity to seizure.

Most workers divide the observational sequence into two or more phases:

Prestimulus behavior (recorded by a few investigators). This involves a description of the variety of actions displayed by the *S* in its home cage and in the test chamber just prior to being exposed to sound. Crude estimates of amount and vigor of activity are frequently included.

Preconvulsion behavior. The *S*'s behavior from the onset of auditory stimulation until the appearance of the violent behavior that characterizes the fit. Data from this period include descriptions of premonitory (i.e., the crouching, shivering, burrowing, etc., that precedes convulsion) and substitute behavior (i.e., the preening, washing, exploring, and attacking the sound source that may appear on nonconvulsion trials). Occasionally estimates of the frequency and duration of the several symptoms are included.

Convulsive behavior. That phase of the *S*'s reaction characterized as wild, violent, uncontrolled, or quasi-uncontrolled. Typically, concern is with the number, and sometimes the duration, of convulsive reactions in a group of animal *S*s or in a single *S* on a series of trials. Reactions are classified as running, subconvulsive, abortive, or disturbed when the animal's reaction does not progress beyond vigorous, apparently blind, running and circling movements (**7, 88, 101**). Full convulsive attacks involve spasms and clonic and/or tonic-clonic episodes. An additional category may be used if the convulsive reaction is so severe as to compel interruption of the test to prevent death or if, as may occur in a sequence of tests, a modified "controlled" pattern appears (**55, 84**).

The most common type of quantification consists of the relative frequency of convulsion, or various types of convulsion, in a group or in individual *S*s for a series of tests (**7**). Such scores may be submitted to probit analysis (**31**). Other forms of scoring have been more recently introduced. Estimates of severity have been derived by assuming that the ratings from "no reactions" through "running" to "tonic-clonic attack" form a linear progression and thus may be summed or averaged (**8**, **88**, **89**). Latency measures and duration measures have also been obtained for the running attacks and the tonic-clonic episode (**7**, **8**, **33**, **45**, **55**, **96**). The mean number of running attacks as well as the relative duration of the pre-epileptoid running have also been useful indices (**8**). No one, however, has yet attempted to devise a chronaxie-like score that incorporates both intensity and duration.

Postconvulsive behavior includes coma and *cerea flexibilitas* or occasionally "epileptoid furor." This infrequent "bouncing response" is usually associated with less severe tonic-clonic attacks. Reports may also include *recovery time* (the duration of the interval between the cessation of convulsive movements and the appearance of "normal" movements [**101**]) or *death time* (the interval from the onset of the stimulus until relaxation of the flattened pinnae [**44**]). Occasionally an appraisal of the *S*'s sensitivity to various forms of further stimulation is made (**101**, **128**).

VARIABLES INFLUENCING SEIZURE SUSCEPTIBILITY

Characteristics of the Seizure Situation

The rigidity of early views concerning the relevance of certain aspects of the test situation for seizure-induction has precluded extensive empirical delineation of their significance. Though a few investigators (**28**, **96**, **103**) have emphasized that sound as both a sufficient and necessary condition has not yet been adequately established, the majority of investigations to date have been conducted on this assumption.

"Crucial" Sensory Modality

Audition. A few studies have dealt with the relative effectiveness of different kinds of auditory stimulation in provoking attacks. Comparisons of simple and complex sounds generally indicate the superiority of noise (bells, jingling keys, etc.) over pure tones (**80**, **104**). This may be due to nonsonic stimulation (vibration of the test chamber) or to the acoustic property of complexity (**31**). When the noise of jingling keys was transmitted through an acoustic system which filtered out frequencies above 6,000 cps, seizures failed to occur; in contrast an effective band was found between 8,000 and 11,000 cps, the upper limit reproduced (**104**). An arrangement of three oscillators, with the high frequency beating against the low, was as effective as keys: the optimum high frequency, 9,500 cps; the low between 50 and 500 cps modulated two or three times per second; the minimum intensity, 20 baryes. Pure-tone experiments indicate the most effective frequency to lie above 8,000 cps. Morin, *et al.* (**104**), working with rats, suggests 9,500 cps as the optimum; McGrael (**80**) using DBA mice, a peak range between 10 and 12 kc. Frings and Frings (**32**), also using mice, report the range of maximum effectiveness to be between 12 and 25 kc. with seizures evokable by frequencies as high as 35 kc. Most investigators have used stimulus intensities of about 100 db. Frings and Frings (**32**) note that with a 10 kc. stimulus, a decrease in intensity from 110 to 95 db results in fewer seizures and longer latencies. Furthermore, greater intensity (median 105 db) is required to initiate than to sustain seizures (95 db) and less to initiate convulsions in subjects with a history of seizure (83 db) than in those convulsing for the first time (92 db) (**31**).

The Morin data on synthetic noise suggest that its superiority may lie in its aperiodicity (**104**). An intermittent pure tone (9,500 cps, 20 baryes) was found to be most effective with three-per-second interruptions, a rate similar to that of certain bioelectric potentials in epilepsy. Prerequisite duration of sound stimulation is also of importance. During a single test, this is most obviously correlated with frequency and intensity, some combinations of these dimensions being more effective than others. But the induction process is probably not a simple monotonic shift toward threshold. Investigators have long known of an upper time limit for the occurrence of the reaction. Frings and Frings (**31**) note that few seizures occur in their mice after one minute. Using the classical priming method of physiology, Fuller and his collaborators (**43, 44, 45**) have shown number of animals convulsing and the mean convulsion latency to vary with duration of the conditioning stimulus and of the conditioning-test stimulation interval. Stimulus priming of 5 to 10 seconds plus an interval of silence up to 60 seconds resulted in a facilitation of seizures. Fifteen seconds of priming plus a 40-second interval produced a reduction in seizure risk. Twenty seconds of priming and 40–60-second intervals yielded better protection. Intervals of 80–100 seconds were associated with a return to normal susceptibility.

Number of previous tests and/or their temporal distribution are other relevant variables. It is common practice to test rats on alternate days in order to allow sufficient time for recovery. Maier, Feldman, and Longhurst (**84**) report animals tested at weekly intervals to have a higher proportion of seizures than those tested daily over a comparable period. Meanwhile, as the number of tests increases, the pattern of dominant response tends to shift, the diffuse reaction being replaced by a controlled forepaw clonus and postconvulsive hyperexcitement (**18, 55, 84**). The greatest number of maximum reactions tends to occur on the first tests (**142**). While latency of running has been variously reported to increase from the beginning to the end of a test series (**96**), and to be high on the first reaction but lowest in the middle of the sequence (**37**), number of successive seizures is regarded by some (**32, 33**) to be a function of the animals' physiological condition rather than the number of previous tests.

Other modalities. Neither of the early explanatory biases—conflict or sensory overflow—require the assumption that auditory stimulation is uniquely necessary to the occurrence of the fit. Thus several experimenters have explored the possible effects of intense excitation of a nonauditory sort.

Intense light. Shafer and Meyer (**125**) exposed rats to intense illumination as well as to combinations of light and sound. No convulsions, either running or tonic-clonic, were provoked by light alone; those obtained with combinations of light and bell occurred no more frequently than with bell alone. Goldberg (**54**) attempted to condition the fit to intense two-per-second flashes of light. In a preconditioning series, light alone elicited no seizure symptoms. In the nonreinforced test trials, several bursts of wild running were noted. The experimenter, however, is hesitant to regard these as conditioned convulsive reactions.

Pain. Several experimenters have commented on the possible significance of the extra-auditory pain pathway. Marx and Jurko (**88**) suggest that irritation and increased tension on the tympanic membrane may reduce the pain threshold and thus facilitate seizures. Shafer and Meyer (**125**), pointing out that the pinna is supplied by the trigeminal nerve, propose that susceptibility may be associated with pinna irritability.

Electric shock has been described as an effective stimulus on several occasions. The present writer has observed a severe epileptoid reaction in each of two rats forced with intense shock at the choice point of a gridded Y maze. Meanwhile, Lazovik and Patton (**75**) report inability to induce convulsions in rats forced with shock to respond to an unsolvable problem arranged for the Lashley jumping stand, even though the subjects were sensitized by a magnesium deficiency. Nor was Coleman (**21**) able to condition the seizure to shock, although he could vary the pattern of reaction by changing the intensity of shock presented with an effective auditory stimulus.

Temperature. Early reports indicated that temperature affected susceptibility although the extent of its influence was confounded by concommitant exercise effects. In a recent experiment Fuller and Rappaport (**42**) found that mice immersed in 5° and 23° C. water immediately prior to test were rendered immune, and convulsive risk in those subjected to 38° C. water significantly reduced. Animals wet after the convulsion began, however, received no protection. The susceptibility of those dried before testing was similarly unaffected. Meanwhile, severity increases with increase in the temperature of the test chamber (**125**).

Proprioception. Speculation has frequently placed proprioception among the "crucial" nonauditory factors. Observed changes in susceptibility coincident with rotation of the animal, restricting its area of movement, binding it and subjecting it to forced exercise antedate the period of this review. Vigorous running and awkward backing and sidling movements are included in all descriptions of the audiogenic attack. Chance's description (**20**) of a strain of mice that convulse when gently shaken sharply contrasts with Fring's (**37**) report that mice may sometimes be revived by such treatment. Two recent studies have dealt with the possible effects of damage to proprioceptive structures upon seizure susceptibility. Finger, Bice, and Day (**28**) found injections of streptomycin induced impairment of righting and motor coordination not seen in control animals treated with dihydrostreptomycin or isotonic saline. No significant differences occurred in the proportion of convulsing rats per treatment group or in the total percentage of convulsions. This supports Marx and Jurko's (**88**) report that labyrinthitis is not a sufficient cause for convulsions. However, no appraisal was made of the extent of vestibular damage and the criteria for evaluating possible behavioral effects were rather crude.

Bevan and Hunt (**8**) obtained no significant differences in the frequency or severity of attacks, nor in the latency for the first running attack, in magnesium-deficient rats tested before and after bilateral cervical dorsal hemisections. However, the operated animals had significantly longer latencies for the epileptoid phase, spent a longer total time and a greater proportion of the interval prior to tonic-clonic attack in running. The tonic-clonic episode, furthermore, appeared to be shorter and less severe in the experimental group. Meanwhile, it is not clear that the experimental lesions were sufficiently extensive or that the dorsal columns in the rat are proprioceptive in function.

Regulation of Free Movement

Interference with the *S*'s movements is a demonstrated variable. Its effects have been variously attributed to shift in proprioceptive inflow, gradual rather than sudden motor outflow, and enhanced conflict. The techniques for effecting interference have varied with the speculative bias of the experimenter. Cain and Naquet (**15**) have corroborated earlier reports that complete immobility through binding affords almost complete seizure protection. Confinement in a small cage or glass cylinder yielded the same results. Cain, *et al.* (**14**) claim similar results when subjects were bound for EEG recording. Griffiths (**61**) reports that rats placed in a Carmichael sock to facilitate the attachment of EEG and EKG electrodes, and tested either in the sock or in a larger box, showed almost no responses. In contrast, a high incidence occurred in rats tamed to the point of permissively allowing the electrodes to be attached; speculation about tameness and neurosis may thus be relevant. The restrained ungentled rats engaged in violent clawing, shaking, biting, and scratching at the electrode leads, the tamed unrestrained showing almost no manipulative behavior at all. Frings and Frings (**31**) report failure to provoke attacks in mice suspended in the test chamber by means of a spring paper-clip, while Rabe (**121**) claims shorter latency, longer convulsions, and quicker death in subjects confined to a small cage inside the tub. Chance (**16**) similarly describes greater responsiveness in amphetamine-treated mice tested in small jars as compared to larger boxes. Michels and Bevan (**96**) meanwhile arranged a situation for altering the size of the test cage without changing the acoustic properties of the auditory stimulus and found no variation in frequency of either running or full convulsions coincident with shift in cage size. Latencies for the first running attacks, however, were longest in the smallest cage. This suggests that under certain circumstances, sensitivity may be directly rather than inversely related to activity level. In line with this are reports of 18 of 29 rats having seizures when tested in a rotating drum (**15**).

Opportunity for Escape

This may be described as a set of conditions that allows the animal *S* to manipulate or otherwise act upon some aspect of the test situation with the result that a convulsion is averted. Conflict-oriented experimenters have interpreted the confinement and restraint situations as ones in which the animal is driven to respond when response is impossible. Griffiths' data, just previously cited, suggest that directed activity is prerequisite to immunity. Maier (**85**) attributes the general ineffectiveness of electric shock in the forced jumping situation to its being a localized stimulus capable of eliciting a specific response and Marx (**89**) suggests that "substitute behavior" is effective because of its pain-reducing influence. If an aspect of the situation draws the *S*'s attention away from the noxious stimulus, enabling him to "go out of the field," it may provide seizure protection. Marx and Van Spankeren (**89**) trained animals to cut off the auditory stimulus by tilting a pole and observed fewer running attacks and tonic-clonic episodes than in an untrained group even with delays in the presentation of the pole up to 64 seconds. Further evidence of control is the predominance of forepaw attack in the trained animals. Goodson and Marx (**55**) confirmed these results with an improved design involving a wheel-turning response. A comparable incidence of reactions occurred in E and C groups when the experimentals were tested without access to the wheel. More recently Chance (**18**) has arranged to introduce intermittently a plasticine hut into the test situation as a "refuge" for the *S*. On the first and second series of tests, introduction of the hut was accompanied by a suppression of the convulsive response persisting for several days even though the hut was removed shortly after the sound began. In a third series a horseshoe of the same dimensions was substituted for the hut without affecting the results. Typically, when the hut was not present, animals displayed searching movements, creeping to the place where the hut had been and relaxing.

Characteristics of the Seizure-Sensitive Animal

The convulsive syndrome is dictated not only by the external conditions that comprise the test situation, but by states of affairs within the animal *S*. These latter may be constitutional factors, specified in physiological or psychological terms, or experimentally produced pretest changes in physiological or psychological status. Statements about individual differences in susceptibility are simply recognitions of this fact. Rat colonies containing inbred stock yield up to about 50 per cent susceptibles. Mouse strains have been developed in which both the convulsive and death risk is high, in which convulsive risk is high but death risk low, and in which both convulsive and death risks are low—but in no case are risks either zero or 100 per cent.

Genetic Differences

Experimenters have generally agreed that such differences are important. Since in most cases, however, it has been impossible to specify the purity of the test stock, or assess possible confounding by differences in nutritional status, age at test, and experimental treatment, the relevant rat data defy confident interpretation. Even with homogeneous mouse stock such considerations as age at test (**37**, **132**) and criteria used to express responsiveness (**40**) have proved of critical importance.

Patterns of reaction. In a 1951 study Frings *et al*. (**37**) reported no strain differences in convulsive pattern among mice used. Vicari (**134**) in the previous year had reported three patterns that varied with age and strain: violent running plus tonic-clonic convulsions and death; less violent running and convulsing with few fatalities; and a static and paralytic reaction with no deaths. In a later study, Frings and Frings (**32**) describe DBA/1 mice as highly susceptible to fatal tonic-clonic reactions, while the responding C57/6's display a long tonic period. There are also differences among the specially developed Frings strains (**30**): one shows frequent clonic, few tonic-clonic reactions; another a high tonic-clonic rate, but few fatalities. The postepileptic phase in rats has been generally one of coma and reduced activity. Chance (**17**), in contrast, describes a terminal saltatory attack in Peromyscus during which the subject bounces with retroverted head, attacking other mice, or models, and sometimes biting the floor. The present writer has observed the same pattern in young Sprague-Dawley rats sensitized by a magnesium-deficient diet, although no tests of attack behavior were made.

Frequency of convulsive reaction. The first experiment comparing mouse strains on responsiveness was conducted by Hall (**64**), who found that 93 per cent of a group of Bar Harbor DBA animals gave violent tonic-clonic reactions, while only 11 per cent of his C57 sample convulsed. Three DBA transplants into C57 mothers all convulsed. About the same time Vicari (**130**) reported similar differences for these strains, and in addition 83 per cent susceptibility in the CE (extremely dilute) strain after 55 days of age, and 28.5 per cent in Albino A at 5 months. Lindsey obtained 80 per cent convulsions in 34-day old DBA's on a single test, 21 per cent in Albino A, 9 per cent in Albino C, 2 per cent in C3H, and none in C57 (**77**). Witt and Hall (**142**) crossed DBA/1 and C57/6 and got 90.5 per cent reactors in the F_1 and 77.3 per cent in the F_2. Backcrossing F_1 with C57 yielded 52.8 per cent, with DBA 90.9 per cent. Finally, a critical backcross of C57 with nonreactors of the C57 backcross gave 25 per cent. Fuller (**39**) crossing DBA/2 with C57 obtained 32 per cent and 58 per cent in the F_1 and F_2 respectively, 92 per cent in DBA B_x and 23 per cent in the C57 B_x. Ginsburg, Miller and, Zamis (**50**) in crosses of DBA/2 and C57/10 got 40 per cent more seizures in the F_1 and F_2 than when DBA/1 was bred to C57/10. Replacing Witt and Hall's four tests with five, and reactor-nonreactor classification with scores of convulsive risk, Fuller, Easler, and Smith (**40**) obtained F_1's with 68 per cent and 79 per cent convulsers, F_2, with 70 per cent, and $F_1 \times$DBA and $F_1 \times$C57 with 95 per cent respectively—all lower, with the exception of $F_1 \times$DBA, than the proportions of Witt and Hall. Later Fuller and Williams (**45**) describe a convulsive risk of 11 per cent for the A/He strain and the following for hybrids: DBA/2×C57, 76 per cent; DBA/2×A/He, 91 per cent A/He×C57, 4.5 per cent. In a comparison of DBA lines 1 and 2 and C57 lines 6 and 10, Miller, *et al.* (**98**) report an 81 per cent incidence for DBA/1, 98 per cent for DBA/2, while both C57's were less than 5 per cent. In F_1, however, hybrids of C57/10 showed a greater incidence than those of C57/6. All F_2's were lower than F_1 with the exception of DBA/1×C57/6 which remained about the same.

Severity of reaction. Little has been done with severity as a scalable parameter. Witt and Hall (**42**) classed reactions into three intensity categories: running; running plus tonic-clonic episodes; running, tonic-clonic attack, and death. If these categories be assumed to represent equally spaced intervals on a three-point scale, a composite severity score may be derived by summing the products of these weights and the relative frequency scores for each category. Witt and Hall's groups rank as follows on this criterion: DBA, very severe; F_1 and DBA B_x, slightly less severe; F_2 severe; C57 B_x, moderately severe; C57 and $C57_x$ nonreactors of C57 B_x, of negligible severity.

Latency of reaction. Strains differ in their readiness to succumb to the seizure-inducing stimulus. Fuller, *et al.* (**40**) present significantly shorter mean convulsion latencies for DBA and $F_1 \times$DBA (35.8 sec. and 38.9 sec). than for F_1 and $F_1 \times$C57 (55.2 sec. and 54.5 sec.). Latencies in the last two groups, and possibly F_2, were bimodally distributed. Bimodality in DBA/2 is more clearly demonstrated in a later study (**44**). A/He is considerably slower to react than

DBA, having a mean latency of 61.8 sec. (**45**); hybrids of A may be slow or fast, depending on the nature of the cross; A×DBA, 35.6 sec.; A×C57, 54.6 sec. In contrast, Frings, Frings, and Kivert's data (**37**) suggest no difference in the latency for the first running attack for DBA, C57, and mongrel Swiss Albino (average times: 33 sec., 31 sec. and 33 sec.). A later paper (**32**) reports average latencies of 26.6 sec., 26.1 sec., and 21.0 sec. for these same strains.

Death risk. Unless suffering from severe nutritional defect, rats only rarely succumb. Not so with mice. Hall (**64**) reported 87 per cent fatalities in his DBA subjects, 7 per cent in his C57 group. Vicari (**130**) has obtained similar results: at 30 days of age 67 per cent of her DBA died; at 35 days fatalities were reduced to 44.4 per cent. In addition, while her CE's, Albino A's, and C_3H's differed in responsiveness, all were free of casualties. A later investigation (**135**) reports the highest DBA mortality to occur in the period 30–39 days. Lindzey (**77**) notes the following strain differences in 34-day old samples: DBA, 75 per cent deaths; Albino A, 1 per cent; C57, C_3H, and Albino C, all zero. Frings' (**37**) mongrel Swiss Albinos show slightly greater inclination toward fatalities than C57. More recently, Frings and Frings (**34, 35**) have developed two strains that are highly responsive but show exceedingly few deaths. Thus susceptibility to seizure and capacity for recovery appear to depend on independent mechanisms. This is further indicated by Fuller, Easler, and Smith's (**40**) failure to discover a significant relationship between latency and fatality and their finding of the same convulsive risk in DBA and backcross of C57 B1×DBA with lower death risk in DBA, as well as by Fuller and Williams' report (**45**) of high convulsive risk but low death risk in the hybrids of A and DBA/2. It is highly probable that C57 would succumb readily if they were less seizure resistant.

Age of first reaction and greatest sensitivity. Mouse strains differ in both the mean age of first convulsion and period of greatest sensitivity. While Vicari reports 78 per cent reactors in DBA at 30 days, and about the same at 35 days (**130**), C57 tests yielded 83 per cent responses at 30 days, and only 8 per cent at 35 days. No reactions appear in Albino A and C_3H before three months. There is also an upper limit for sensitivity. Vicari (**135**) indicates this to be 80 days for DBA. Frings, Frings, and Kivert (**37**) report that DBA's convulse first at 18.7 days, C57's at 20.5 days, and mongrel Swiss Albinos at 20 days, with none of the between-groups differences significant. All three strains are said to be most sensitive before 30 days: DBA at 22, C57 at 24, and Albinos at 24 days. The DBA datum is at variance with Vicari's estimate, and probably reflects a difference in test procedure. It is clear that C57's are responsive if tested prior to 30 days. Dice and Barto's (**24**) discovery that Peromyscus, with an auditory range of from 500 to 95,000 cps, shows a decrease in auditory sensitivity with age may clarify these seizure results. It is also apparent that a certain level of neuromuscular development is prerequisite to convulsions. Frings and Frings (**32**) describe the movements of 12–13 day old animals as premonitory, yet they never culminate in an attack.

Strain, sex, and sensitivity. There is little among the rat data to indicate sex differences in susceptibility. The earlier mouse studies prompt the same conclusion (**131, 142**). But Vicari (**132**) has observed that endocrine therapy with several strains reduced seizure fatalities 40 per cent in males and 90 per cent in females. The protective effect of glutamic acid is also greatest for DBA males (**41**, **47**). In line with this is Frings, Frings, and Kiverts' (**37**) conclusion that males are more fragile than females (14 male of 23 deaths in the DBA, two males of two in the C57 and three males of seven in the Swiss Albino sample). All of these studies extend the findings of Ginsburg and Hovda (**48**) with DBA animals. Meanwhile, the Fuller, Easler, and Smith (**40**) data also suggest that sex is related to susceptibility, its influence being controlled by the genotype. In highly susceptible or relatively immune stocks sex makes little difference; in stocks of moderate susceptibility it is an effective variable. When responses of offspring of nine possible combinations of DBA, C57 and their F_1 were evaluated, using first trial seizure risk as a criterion, males had a higher percentage of seizures in five comparisons, females in two and in two there was no difference. The largest differences were in the $F_1 \times C57$ back cross and in the F_1 hybrids. Additional support is provided by the higher incidence in DBA/2×C47/10 males as contrasted to DBA/2×C57/6 males (**98**).

The genetic mechanism of susceptibility. Speculation about genetic transmission in the rat has centered about three modes: single dominant gene, a pair of genes, and multifactor determination. The differences in theorizing reflect differences in empirical data that have been attributed to the use of impure strains (**65**) and inadequate control of other pertinent variables (**27**). The picture on homogeneous mouse strains also lacks complete unequivocality. From the proportions of *reactors* (animals giving at least one convulsion in a series of four tests), obtained with DBA/1, C57/6 and the offspring of crossing these lines, Witt and Hall (**142**) have concluded that mouse susceptibility is governed by a single dominant gene. However, their prediction for the critical back cross (C57×nonreactors of $F_1 \times C57$) was not confirmed. This prompts the recognition of "minor modifying factors." Ginsburg and Hovda (**48**) hold the Witt and Hall conclusion, but point out that susceptibility varies with age. In a later experiment involving crosses of DBA/2 and C57/10 Ginsburg, *et al.* (**50**) obtained proportions in F_1 intermediate between those of the parent strains, and in F_2 three-fourths that of F_1, and propose the presence of two or more nondominant factors in addition to a "convulsing" gene. Frings, Frings, and Kivert (**37**) reconcile their report of maximum sensitivity in DBA and C57 during the 20-day period with the Witt and Hall and the Vicari results for 30-day animals by suggesting that the two strains differ with respect to a single gene pair, the recessive allele producing only the early sensitivity, the dominant being associated with both the 20- and 30-day susceptibility.

Fuller, Easler, and Smith (**40**), noting the failure of the Witt and Hall (**142**) critical backcross and the divergence of results (**50**, **142**) with different sublines of DBA and C57, tested the Witt and Hall hypothesis by reproducing their test procedure (except that seizure risk on first test rather than reactor-nonreactor classification after four tests served as the susceptibility index) with C57/6 and DBA/2. Two F_1 samples were intermediate between the parent strains, susceptibility appearing to decrease in accordance with the proportion of DBA genes in each group. Contrary to the single gene hypothesis, however, the proportions of F_2 and F_1 convulsing in the first trial were significantly less than expected. Otherwise obtained convulsion frequencies did not defy the assumption that a single gene produces high susceptibility in homozygous material but only moderate sensitivity in heterozygous. The possibility of multigene determination meanwhile is suggested by almost identical seizure risks in an F_3 of nonconvulsing F_2 *S*s and the $F_1 \times C57$. The variability in the F_1 indicates susceptibility is a characteristic that fluctuates within a wide range about the average convulsive risk, nongenetic factors perhaps determining position within this range. This would be most important in heterozygous animals whose mean position on the susceptibility dimension would be closer to the threshold than that of homozygous strains.

Though susceptibility and the capacity for recovery are both inherited, they do not appear to be genetically linked. This last fact and the priming data are elaborated by Fuller and Williams (**45**) and Fuller and Smith (**44**) into a kinetic theory of seizures: the occurrence of a convulsive reaction depends upon the balance achieved between a gene-controlled, rapidly developing excitatory and a more slowly aroused accommodatory process. Strain differences in convulsive risk and latency reflect differences in the rates at which these two mechanisms reach critical functional level.

Age and Sex Differences

Little more need be added concerning these factors. Maier, Feldman, and Longhurst's (**84**) study of changes in seizure pattern segregated age from practice effects by testing animals daily and weekly over the same age range. They report that the rat's reaction on the first test is prognostic of his responsiveness on later tests. They also found greater sensitivity in the animals tested weekly. Running fits and full epileptoid fits occurred predominantly in the early weeks of testing, a modified, controlled type of seizure frequently followed by hypertenseness becoming the predominant later response.

The effects of susceptibility produced by a 10 per cent lactose supplement differ for male and female rats (**86**). In males a decrease in the number of seizure-prone animals occurs without effecting the frequency of response in the reactors; in females the frequency of reactions decreases with no change in number of convulsions. Since the treated group displayed more epileptoid reactions than the controls, it was concluded that lactose enhances irritability in seizure-prone *S*s, this change being more general in females. The smaller number of male reactors was attributed to a sex difference in the optimum proportion of lactose. However, since there are no data to indicate what the prelactose sensitivity of the experimental animals might have been, it is not improbable that the sex-lactose relationship is confounded by other variables.

Vicari, Tracy, and Jongbloed (**138**) have found that the administration of testosterone, pregnenalone, estradiol, cortisone, and cortin all alleviate convulsion and death to one degree or another. So also does glucose, epinephrine, and thiouracil. Ginsburg and Roberts (**51**) report sex differences in the effects of a variety of drugs upon the incidence of convulsions. Injections of glutamine aggravated seizures in male DBA mice, not females; methionine sulfoxide, and α-methyl glutamic acid had a greater influence in females; γ-glutamyl hydrazide depressed sensitivity in females, enhanced it in males; dinitrophenol depressed activity in males; and cortisone and γ-amino-butyric acid enhanced convulsions in females.

Differences in Physiological Status

Physiological changes during convulsion. Hofeld (**69**) describes a sharp increase in BMR during the audiogenic attack, with a drop to subnormal level in the comatose state following attack—all expected changes. Ginsburg and Hovda (**48**) claim a decrease centrally in free acetylcholine and a complementary peripheral increase. While there was no difference in free acetylcholine in brain extracts of untested DBA and C57 animals, the submaxillary gland extracts of the former contained significantly less than those of the latter. Abood and Gerard (**1**) report defective phosphorus metabolism during the susceptible period: ATPase, alkaline phosphotase, and P/O ratio were all below normal as was liver glycogen, and there was a slow turnover of phosphorylated intermediates. Brains were lighter in susceptible strains. Cain, Mercier, and Corriol (**14**) describe the large waves in EEG records made during tests reported earlier by Lindsley, Finger, and Henry (**76**). These they regard as artifacts of vigorous movement. No other consistent change occurred save an occasional diminution in the large spikes during the passive phase. They thus conclude that audiogenic seizure lacks the electrographic characteristics of experimental epilepsy. Griffiths (**61**) more recently has obtained both slower heart and slower brain tracings during the attack, which, in contrast, he attributes to a reduction in movement (manipulative movement).

Prenatal environment and responsiveness. Hall (**64**) describes three convulsions and two deaths in three DBA fetal transplants into C57 mothers when tested at 28 days and suggests the importance of heredity as contrasted to prenatal environment in determining the constitutional bent of certain strains for convulsions. Ginsburg and Hovda (**48**), while confirming Hall on convulsiveness, found fewer deaths in DBA's born to and reared by C57 mothers than in nontransplants.

Dietary variation and responsiveness. Most of the studies in this category deal with the influence of the B vitamins upon susceptibility. Patton (**109**) has pointed out that while they do not explain convulsive activity, they demonstrate how the sensitivity of a seizure-prone, or potentially prone, subject may be intensified or alleviated. They thus offer a technique for a more precise experimental control over the dependent variable. Patton (**113**) extending his previous findings on the B complex, reports that young animals maintained on a basal diet including thiamin, riboflavin, pantothenic acid, choline, and pyridoxine display a high incidence of convulsions. The addition of five more vitamins: nicotinic acid, para-aminobenzoic acid, biotin, folic acid, and inositol; rice polish concentrate; ventriculin; or liver extract resulted in a marked reduction in convulsiveness. Less striking results occurred with supplements of brewer's yeast or with the feeding of Purina Chow. The animals on the basal diet were otherwise apparently healthy. Sensitivity can also be reduced by supplements of pyridoxin, if a deficiency of pyridoxin has been induced, this latter requiring feeding of the deficient diet from birth (**120**). In a cleverly conceived experiment, Griffiths (**60**) tested for susceptibility during two-week periods when his *S*s were maintained on a stock diet and a self-selection diet. Three highly susceptible animals and three of low sensitivity ceased having fits while on the self-selection diet; five of these took in atypically large amounts of thiamine hydrochloride, the sixth, of magnesium chloride.

Several studies have dealt with convulsions in rats on sugar-supplemented diets. The Maier, Longhurst, and Ellen (**86**) study, already referred to, substituted 10 per cent lactose for cornstarch in the pregnant mother's diet and that of the offspring for nine months and resulted in an increase in sensitivity in the latter. However, according to Benedict (**5**), a 20 per cent lactose supplement from weaning until three or four months of age has no effect upon the frequency of reactions. The same is true for 20 per cent supplements of dextrose. Ginsburg and Roberts (**51**) also report supplements of dextrose to have effect with DBA/1. On the other hand, Vicari, Tracy, and Jongbloed (**138**) claim 5 and 10 per cent supplements of glucose to have a protective effect.

Bevan, Hard, and Seal (**7**) induced shorter response latency and greater frequency and intensity of seizure in young rats by feeding a lysine-deficient diet. A return to Purina Chow restored normal sensitivity. Animals maintained on the experimental diet also show high and more variable wheel activity, poorer maze performance, greater irritability, reduced food and water intake, loss of body weight, hypoglycemia, low hemoglobin level and red-cell count, lower total serum protein, and markedly reduced clotting times.

Hamilton and Maher (**67**) included a series of seizure tests in an experiment on the effects of glutamic acid supplements upon rat behavior. The number of seizure-prone animals, however, proved too few to warrant any conclusion concerning susceptibility. Ginsburg, *et al.* (**53**) found that while supplements of glutamic acid had no effect upon seizure incidence in DBA mice, fatalities decreased. This protective action they attribute to the fact that glutamic acid is a nonessential amino acid. Since it is derivable from α-ketoglutaric acid, any excess of it may result in a higher cell-concentration of α-ketoglutaric and consequently an increase in energy output via the tricarboxylic acid cycle. The animal thus has energy enough to prevent severe spasms and/or to survive respiratory arrest. Later Ginsburg and Roberts (**51**) found not only reduction in deaths but also in seizures after 1+glutamic acid supplements. Fuller and Ginsburg (**41**) rejected the possibility that the protective effect of glutamic acid is due to adrenergic action after finding its persistence following adrenalectomy and adrenalectomy with DCA implants. The protective effect of glutamic acid in females may be further enhanced by using an interrupted rather than a continuous stimulation schedule (**47**). Other metabolically active substitutes also have a palliative effect; lactic acid, succinic acid, and γ-glutamyl hydrazide (**51**). Others, however, enhanced sensitivity or exerted no effect at all: malic acid, α-ketoglutaric acid, sodium pyruvate, isomolar injections of glutamine, methionine sulfoxide, α-methyl glutamic acid, and γ-amino butyric acid increased responsiveness; alanine, aspartic acid, fumaric acid, and sodium oxalacetate had no observable effect.

In the early part of the period covered by this review the effects of an induced magnesium deficiency were still under consideration. Lazovik and Patton (**75**) indicated that initially insensitive rats could be made to convulse after maintenance on a low-magnesium diet and returned to normal insensitivity by magnesium sulfate supplements. The same was found to be true of hamsters (**112**), which convulsed even after subparalytic doses of curare. Both wild and domestic rats reacted after a magnesium-deficient regimen, although the wild were the more resistant and showed no fatalities (**57**).

Hormone and drug effects. Since the audiogenic seizure has been thought of in terms of emotional stress and emergency state, a vigorous interest has been maintained in the relation of drug and hormone balance to seizure sensitivity. Nicotine facilitates convulsions (**38**). Meanwhile, two studies (**48, 51**) indicate that subshock doses of insulin and two (**23, 56**) that alcohol have a protective effect. It has been suggested that every agent that reduces metabolic rate will protect from seizure. Thiouracil prevents death in DBA and reduces the severity of convulsions (**136**). Reduced body temperature is associated with immunity. This, in part at least, may be related to the epinephrine function. This should mean that overactivity of the adrenal gland would be associated with hypersensitivity. However, if the audiogenic seizure be considered an emergency state, increased output should have a desensitizing function, since it is associated with energy mobilization. Meanwhile, should the convulsion be a symptom of autonomic imbalance, any agent, sympathetic or parasympathetic, which restores that balance should have a protective effect. Three papers (**12, 90, 138**) describe the results of injections of adrenalin as alleviating. The same is true for certain sympatheticomimetic substances: extract of genet (**12**), ephedrine (**90**), Benzedrine (**90**), ergotamine (**93**), and yohimbine (**93**). Mercier (**93**) is inclined to regard these drugs, particularly the last two, to effect sensitivity through a hypotensive action. In contrast, Chance (**16**) finds enhanced sensitivity in mice after doses of ephedrine and Benzedrine. This inconsistency is unexplainable unless it may be attributed to a difference in strength of dose or to the possibility that one worker used insensitive, the other sensitive animals, the drugs having no effect in either case.

Griffiths (**59**), attacking the hypothesis that lethargy and inactivity are associated with immunity, tested animals given a one-stage adrenalectomy. No differences in seizure incidence were found to obtain between operated and control animals, either emotional or nonemotional, as indicated by the Hall open-field test. Similar results occurred when Griffiths and Cohen (**62**) induced lethargy with morphine sulfate.

Yeakel, *et al.* (**143**) exposed rats to air blast for daily five-minute periods over the course of a year and produced an increase in blood pressure, and possibly an increase in adrenal weight, not found in unexposed rats. In a follow-up experiment (**79**) animals with elevated blood pressure were subjected to adrenalectomy and replacement therapy with the expectation that if hypertension had resulted from adrenocortical overactivity, blood pressure would return only to normal from the postoperative low; if not, replacement therapy would effect a return to the hypertensive state. The average initial systolic pressure for sound-exposed animals was 125; during the last six months of stimulation, 162; after surgery, 108; after replacement therapy, 125,—prompting the conclusion that air blast constitutes a chronic alarm stimulus and high blood pressure represents a symptom of hyperactivity of the adrenal cortex and the failure of the organism to adapt. In contrast, Ginsburg and Roberts (**51**) report that cortisone raises seizure and death incidence in DBA/1. Hurder and Sanders (**70**, **71**) failed to induce an anticipated change in incidence with injections of ACTH, although the susceptible-treated animals were found to have heavier adrenal glands than the unsusceptible-treated. Finally, Vicari, *et al.* (**138**) described a slight protective effect for cortisone and cortin that is in line with adaptation syndrome expectation. They also found other steroids (testosterone, pregnenolone, and estradiol) to be more effective than cortisone and cortin in reducing both incidence and fatalities.

Because of our present knowledge concerning the action of acetylcholine in autonomic ganglia, injections of this substance in appropriate form may be expected to influence—possibly enhance—sensitivity. Both acetylcholine and Eserine are reported to reduce incidence and fatality (**49**). Prostigmine and atropine were found to reduce deaths. In addition, there is a central decrease with a peripheral increase of free acetylcholine during seizures in DBA animals (**48**).

Since Tridione has been found to alleviate human *petit mal* attacks, as well as experimentally induced drug and electroshock seizures, its effect upon audiogenic convulsions has been explored by Adams and Griffiths (**2**). When given intraperitoneally to a small number of animals in doses comparable to that for the *petit mal* attack (3.7 mg/kg of 1 per cent solution), no effect upon the running attack was observed. When the dosage was increased to 10.8 mg/kg, incidence showed a marked drop. In no case did the drug appear to paralyze or effect confusion in the animals. Mercier and Cain (**94**) report the protective dose to be 300 mg/kg in 5 per cent aqueous solution of urethane given subcutaneously. This is considerably higher than the critical doses reported by these same workers (**13**) for diphenylhydantoine (20 mg/kg) and phenobarbital (3 mg/kg). These last are not toxic although that for diphenylhydantoine is about four times the critical dose for human beings. In their Tridione study, subjects became lethargic with doses of 20 mg/kg and suffered paralysis with doses of 40 mg/kg. Because of this protective effect of Tridione they classify audiogenic seizures as epilepsy. Adams and Griffiths (**2**), because of the difference in critical doses for rat and man, prefer to remain uncommitted on this point. Mercier (**91**) later investigated the effect upon seizures of introducing bromine into the phenobarbital molecule as well as the effect of bromine itself. Comparison of equally strong doses of phenobarbital and metabromogardenal indicated the latter to be slightly more protective. Doses of sodium bromide varying from 50 to 600 mg/kg and giving from 20 to 50 per cent protection were never as effective as metabromogardenal. Thus in the case of this last the barbituric nucleus is predominant. In another study (**92**) the effect of substituting an ethyl for a phenyl radical was explored. Diethylmalonylurea (barbital) was given in dosages comparable to the critical dose of phenobarbital and found to be less effective and less reliable. With double the critical dose, treated animals were unresponsive on about one third of the trials. However, since a complete epileptic attack was never induced, barbital was thought to have a sedative effect. Derivatives of the phenyl 5 group were also replaced by anisyl and piperyl radicals to give dianisyl—5.5-hydantoin and dipiperyl—5.5-hydantoin, but both substances appeared to be without anticonvulsant action. Thus the anticonvulsant effect of hydantoin derivatives depends on the integrity of phenyl 5.

Much of the work on hormones has involved nonphysiological dosages. Thus, while the results may be of pharmacological interest, they add little to our understudy of the relation of normal hormone balance to sensitivity.

Middle ear infection and convulsive sensitivity. In 1947 Patton (**111**) suggested that because of its common occurrence in albino Norway stock, purulent otitis might be considered to be a contributing factor to seizure sensitivity. While only two among a group of male and female breeders, 85 per cent of which were infected, showed signs of responsiveness, a high relation between infection and susceptibility appeared to exist among his young animals: about three-fourths were both infected and susceptible; only a small percentage were infected but nonsusceptible; none free of the disease had a convulsion. Kenshalo and Kryter (**74**) attempted to clarify this infection-convulsion relationship by using white noise as the seizure-provoking stimulus and a more sensitive infection criterion. After a series of auditory tests, the bulla was swabbed and a brain-heart media, inoculated with its contents, incubated for 48 hours at 37° C. The culture was then smeared on a blood-agar base and incubated for an additional 48 hours, each colony being finally stained with gram stain. Other cultures were made in nutrient media containing *d*-mannose. The criteria for infection were gram stain reaction, acid reaction to *d*-mannose, and microscopic appearance. Of 71 animals, 46 were found to be infected in both ears, 10 in one ear, and 15 healthy. Of the "two-ear" group, 85 per cent were reactors (one or more convulsions); of the "one-ear," 90 per cent; of the "no-ear," 80 per cent. Kenshalo and Kryter thus conclude that infection is not a necessary condition to convulsion, but may enhance susceptibility since the "two-ear" group convulsed on 81.5 per cent of the trials in contrast to the "no-ear" 36.3 per cent. Other data also indicate the nonessential role of infection (**90**). Of 25 DBA's subjected to histological examination after seizure test only three showed massive and one mild infection. Among 12 C57 animals, there were two massive and two mild infections. Marx and Jurko (**88**) doubting that Patton's infected rats were purulent, tested 82 animals, using several measures of sensitivity. Those found to be bilaterally diseased had displayed a greater proportion of convulsions than those unilaterally infected. Of 17 animals classed as insensitive, only two had both ears infected; of 23 classed as most sensitive, 17 had both ears infected. Infections, furthermore, were predominantly of the secretory type. The fact that 13 of 17 insensitives were diseased while 4 of 8 normals were sensitive reinforces the conclusion that middle-ear infection is not necessary to seizures. Marx and Jurko (**88**) suggest that infection exerts its influence by bacterial invasion of the inner-ear or brain or by lowering the threshold for extra-auditory pain. Frings and Frings (**29**) checked the possible relationship of infection to seizure in three strains of mice: DBA/1, C57/6 and mongrel Swiss Albino. While animals giving no seizure responses for at least 10 days prior to examination were free of infection, one of those dying in convulsion had otitis. Thus infection was judged to be unimportant for convulsions in these animals. In contrast, Patton and Zabarenko (**118**) in an unpublished experiment involving 355 young rats, found the great majority of their susceptible animals to be infected.

Marx and Chambers (**87**) set about to clarify the role of otitis media by examining for possible changes in sensitivity subsequent to experimentally inducing pathological state. To accomplish this last, a plastic, butyl methacrylate polymer, dissolved in acetone was used to occlude one Eustachian tube in each experimental animal. All but two of the experimental animals showed suppurative infection, these two displaying a secretory disorder; additional infection was found in the untreated ears of six experimentals and four controls. Only three animals in the entire population, however, had convulsions during the study. Thus it was concluded that suppurative infection is not associated with susceptibility. These data, of course, provide no test of the possible importance of secretory infection, nor rule out the conclusion that infection enhances sensitivity in already sensitive animals.

The nervous system and audiogenic seizures. Clinical studies of hydration of the cerebral tissue indicate an enhancement of frequency and severity of human epileptic attacks. The same relation has been found to hold for electroshock and certain drug convulsions in laboratory animals. Auditory stimulation four hours after peritoneal injections of isotonic glucose (100 cc/kg), however, was ineffective in producing a single reaction in insensitive rats (**95**). Of 29 highly sensitive animals, only three gave convulsions four or five hours after treatment. This protective effect, furthermore, appeared to parallel the development of hydration, since two animals of another sample reacted 30 minutes and one hour after injection, none at intervals of three and four hours. When animals were tested 24 hours after injection, only 55 per cent appeared to have regained their pretreatment sensitivity. From these data it was suggested that the mechanism is not the same in audiogenic and electroshock and cardiazol convulsions, but is also not due to conflict.

Brdar and Moyer (**11**) sectioned the IXth and Xth nerves in DBA×C57 hybrids and tested for seizure sensitivity on four consecutive days after operation. Although this cross is reported to be 95 per cent sensitive, none of the 19 glossopharyngeal animals gave signs of convulsing. The sole survivor of vagus sectioning was also immune. Meanwhile, of over a dozen animals subjected to sectioning of the XIth nerve and immune during the regular sequence of postoperative tests, a considerable number succumbed in tests administered on the sixth operative day.

Of the remaining papers to be summarized in this section, one deals with the effect of frontal damage upon seizures; one with destruction of the motor cortex, and two with interruption of afferent pathways. In line with the early report of Weiner and Morgan (**139**), Longhurst (**78**) describes a decrease in both frequency of seizure and number of susceptibles following partial frontal ablation. Seven animals with unilateral damage to the depth of the basal ganglia showed temporary cessation of convulsions, two with deep bilateral injury, apparently permanent immunity. Seventeen with superficial damage showed no postoperative change in sensitivity. Longhurst suggests that deep lesions were effective because they interrupted the motor processes distributing directly to the putamen. Patton (**112**) has speculated on the importance of the midbrain for seizures, since he found subparalytic doses of curare did not prevent attacks in animals rendered more sensitive by restricting magnesium intake. The enhanced sensitivity with cortical ablation observed by Beach and Weaver (**4**) might be explained in terms of cortical release, although with removal of 90 per cent of the cerebrum one might also expect damage to subcortical centers. Further testimony to the importance of subcortical centers is found in Morin and Cain's (**102, 103**) reports that attacks can be produced after destruction in the motor cortex and degeneration in the corticospinal tracts. Postconvulsive stupor and *cerea flexibilitas* are said to be associated with increased thresholds for certain types of stimuli and the abolition of postural corrective and cortical static reactions with the preservation of the reflexes of support and balance.

Two studies have dealt with the possible importance of propioceptive inflow for sensitivity. Finger, Bice, and Day (**28**) produced vestibular dysfunction with a long-term regimen of streptomycin injections (**134**, two per day with doses ranging from .1 gm/kg) but failed to induce any difference in the number of animals convulsing or the frequency of seizures as compared to controls given either dihydrostreptomycin or an isotonic saline. Bevan and Hunt (**8**) sectioned the dorsal columns bilaterally at the cervical level and failed to produce changes in the number or severity of convulsions or differences in latencies for the first running attack. They did obtain, however, significantly greater latencies for the epileptoid phase and significantly longer running time in the operated animals. These last also displayed a greater number of running attacks and spent a greater proportion of the interval between onset of the running and tonic-clonic episode in running. This suggests that kinesthetic excitation, assuming it to be relayed via the dorsal roots in the rat, is relevant for seizure induction. The rather limited effect of surgery could be due to the limited functional significance of this modality, to the possibility that proprioceptive inflow follows another route in the rat, or to inadequate experimental lesions.

Psychological Status

The widespread appeal of the Morgan interpretation of audiogenic seizures has resulted in little consideration, except by Maier and his students, of the possible significance of psychologically specified variables in the etiology of these convulsions. Lazovik and Patton (**75**) for example, point out that to 1947 there was no convincing evidence that conflict produced the attack. To clarify any possible suspicion of its significance, they created what they regarded to be a condition of maximum favorability for its operation: enhanced sensitivity due to magnesium deficiency. In an initial experiment, one group of deficient animals were given 10 trials per day in the jumping situation used by Maier to induce conflict and one daily sound test, a second receiving only the sound tests. Jumping was forced with electric shock. While the members of the first group typically developed resistance to jumping, crouched, failed to explore the jumping platform, and even squealed in the absence of shock, no seizures or prodromal signs were precipitated. Sound-test sensitivity was reduced with dietary-replacement therapy. A second experiment of the same general sort yielded the same results.

Maier and Longhurst (**85**) have refused to regard these experiments as a crucial test of the conflict hypothesis, maintaining that electric shock is a negative stimulus which the animal can localize and thus to which it can react adaptively. Consequently the "trapping effect" necessary for convulsion is not effectively induced. Their position is reinforced by data from a two-group experiment in which rats were trained in a discrimination task and then forced with an air blast to jump to the negative card. Controls, in contrast, received no discrimination training, but were forced to jump in the one-card test situation. The controls, during the course of testing, received, on the average, an air blast of slightly longer duration; however, forced jumping produced seizures in only 7.6 per cent of all tests as contrasted to 20.3 per cent for the experimentals. Similarly 37.8 per cent of the experimentals had convulsions in contrast to 9.1 per cent of the controls, the average number of seizures per animal (1.60) in the first group being significantly greater than that in the second group (.44). While this experiment cannot be regarded as a crucial test of the conflict hypothesis, it is regarded as a strong indication of the basic nature of the conflict interpretation. In a footnote in his 1949 book (**83**), Maier states that Morgan has withdrawn from the auditory reflex-conflict controversy on the basis of these last data.

Among currently active experimenters, Rabe (**121**) uses the sound test as a technique for inducing cumulative frustration. Observations of qualitative changes in the animals' test-cage behavior, and occasionally home-cage behavior, after a series of exposures to sound is consonant with Maier's view. A tractable animal, who initially explores the test-cage, may come to resist handling or removal from the home-cage, squeals when placed in the test chamber, cringes, shivers, and acts as if "it expects something unpleasant to happen" (**37**). Also consonant with a psychological interpretation are Chance and Yaxley's (**19**) and Zener's (**44**) demonstrations of a stage of controlled hyperactivity in the seizure sequence. Slow-motion pictures made by the latter indicate that in the early stages of running, the reacting animal, upon approaching obstacles or the sides of the test-box, turns its head so as to take its collision on a shoulder.

Lindzey's study utilizing five strains of mice (**77**), though it indicates a positive relationship between emotionality and susceptibility, in contrast to previous reports for Hall's emotional and nonemotional rats, does not find the two traits to covary consistently in the various strain groups. Meanwhile, Griffiths has steadfastly regarded constitutional differences in temperament to be associated with differences in sensitivity. He reports (**57**) that wild rats on a magnesium-deficient diet are more resistant to convulsions. He points out that tame animals usually have smaller adrenals than wild; and that wild Norways, less susceptible than wild Alexandrines, are highly pugnacious and fierce. His adrenalectomy and morphine sulphate experiments (**59, 62**) are further attempts to demonstrate a relation between temperament and susceptibility.

Griffiths and Stringer (**63**) attempted to induce differences in emotionality by subjecting young animals to intense stimulation (loud noises, rapid rotation, extremes of temperature, electric shock) and checking for possible effects on seizure-sensitivity in adulthood. No differences were demonstrated between uting directly to the putamen. Patton (**112**) has speculated on the importance of the midbrain for seizures, since he found subparalytic doses of curare did not prevent attacks in animals rendered more sensitive by restricting magnesium intake. The enhanced sensitivity with cortical ablation observed by Beach and Weaver (**4**) might be explained in terms of cortical release, although with removal of 90 per cent of the cerebrum one might also expect damage to subcortical centers. Further testimony to the importance of subcortical centers is found in Morin and Cain's (**102, 103**) reports that attacks can be produced after destruction in the motor cortex and degeneration in the corticospinal tracts. Postconvulsive stupor and *cerea flexibilitas* are said to be associated with increased thresholds for certain types of stimuli and the abolition of postural corrective and cortical static reactions with the preservation of the reflexes of support and balance.

Two studies have dealt with the possible importance of propioceptive inflow for sensitivity. Finger, Bice, and Day (**28**) produced vestibular dysfunction with a long-term regimen of streptomycin injections (134, two per day with doses ranging from .1 gm/kg) but failed to induce any difference in the number of animals convulsing or the frequency of seizures as compared to controls given either dihydrostreptomycin or an isotonic saline. Bevan and Hunt (**8**) sectioned the dorsal columns bilaterally at the cervical level and failed to produce changes in the number or severity of convulsions or differences in latencies for the first running attack. They did obtain, however, significantly greater latencies for the epileptoid phase and significantly longer running time in the operated animals. These last also displayed a greater number of running attacks and spent a greater proportion of the interval between onset of the running and tonic-clonic episode in running. This suggests that kinesthetic excitation, assuming it to be relayed via the dorsal roots in the rat, is relevant for seizure induction. The rather limited effect of surgery could be due to the limited functional significance of this modality, to the possibility that proprioceptive inflow follows another route in the rat, or to inadequate experimental lesions.

Psychological Status

The widespread appeal of the Morgan interpretation of audiogenic seizures has resulted in little consideration, except by Maier and his students, of the possible significance of psychologically specified variables in the etiology of these convulsions. Lazovik and Patton (**75**) for example, point out that to 1947 there was no convincing evidence that conflict produced the attack. To clarify any possible suspicion of its significance, they created what they regarded to be a condition of maximum favorability for its operation: enhanced sensitivity due to magnesium deficiency. In an initial experiment, one group of deficient animals were given 10 trials per day in the jumping situation used by Maier to induce conflict and one daily sound test, a second receiving only the sound tests. Jumping was forced with electric shock. While the members of the first group typically developed resistance to jumping, crouched, failed to explore the jumping platform, and even squealed in the absence of shock, no seizures or prodromal signs were precipitated. Sound-test sensitivity was reduced with dietary-replacement therapy. A second experiment of the same general sort yielded the same results.

Maier and Longhurst (**85**) have refused to regard these experiments as a crucial test of the conflict hypothesis, maintaining that electric shock is a negative stimulus which the animal can localize and thus to which it can react adaptively. Consequently the "trapping effect" necessary for convulsion is not effectively induced. Their position is reinforced by data from a two-group experiment in which rats were trained in a discrimination task and then forced with an air blast to jump to the negative card. Controls, in contrast, received no discrimination training, but were forced to jump in the one-card test situation. The controls, during the course of testing, received, on the average, an air blast of slightly longer duration; however, forced jumping produced seizures in only 7.6 per cent of all tests as contrasted to 20.3 per cent for the experimentals. Similarly 37.8 per cent of the experimentals had convulsions in contrast to 9.1 per cent of the controls, the average number of seizures per animal (1.60) in the first group being significantly greater than that in the second group (.44). While this experiment cannot be regarded as a crucial test of the conflict hypothesis, it is regarded as a strong indication of the basic nature of the conflict interpretation. In a footnote in his 1949 book (**83**), Maier states that Morgan has withdrawn from the auditory reflex-conflict controversy on the basis of these last data.

Among currently active experimenters, Rabe (**121**) uses the sound test as a technique for inducing cumulative frustration. Observations of qualitative changes in the animals' test-cage behavior, and occasionally home-cage behavior, after a series of exposures to sound is consonant with Maier's view. A tractable animal, who initially explores the test-cage, may come to resist handling or removal from the home-cage, squeals when placed in the test chamber, cringes, shivers, and acts as if "it expects something unpleasant to happen" (**37**). Also consonant with a psychological interpretation are Chance and Yaxley's (**19**) and Zener's (**44**) demonstrations of a stage of controlled hyperactivity in the seizure sequence. Slow-motion pictures made by the latter indicate that in the early stages of running, the reacting animal, upon approaching obstacles or the sides of the test-box, turns its head so as to take its collision on a shoulder.

Lindzey's study utilizing five strains of mice (**77**), though it indicates a positive relationship between emotionality and susceptibility, in contrast to previous reports for Hall's emotional and nonemotional rats, does not find the two traits to covary consistently in the various strain groups. Meanwhile, Griffiths has steadfastly regarded constitutional differences in temperament to be associated with differences in sensitivity. He reports (**57**) that wild rats on a magnesium-deficient diet are more resistant to convulsions. He points out that tame animals usually have smaller adrenals than wild; and that wild Norways, less susceptible than wild Alexandrines, are highly pugnacious and fierce. His adrenalectomy and morphine sulphate experiments (**59, 62**) are further attempts to demonstrate a relation between temperament and susceptibility.

Griffiths and Stringer (**63**) attempted to induce differences in emotionality by subjecting young animals to intense stimulation (loud noises, rapid rotation, extremes of temperature, electric shock) and checking for possible effects on seizure-sensitivity in adulthood. No differences were demonstrated between the traumatized and the control animals. Rearing the animals in what the experimenters (**9**) chose to regard as different environments (6 ft.×9 ft. sheet-metal-lined room with sawdust floor; 2 in.×4 in. hardware-cloth squeeze boxes in the large room; 14 in.×9 in. individual cages) also failed to affect susceptibility. Tests in which rat *S*s manipulated some aspect of the environment (**61**) resulted in immunity. When pretest experience was such as to reduce such manipulation, convulsions occurred.

Attempts to induce changes in sensitivity through specific training procedures have had one or the other of two aims: to provoke convulsions with initially ineffective stimuli, or to reduce the efficiency of stimulation already classifiable as seizure-inducing. Two experiments of the first sort have been unsuccessful. Coleman (**21**), investigating the convulsion as a fear reaction, tried to condition the convulsion to electric shock. While he failed to obtain a single convulsion in eight *S*s, seven of these animals came quickly to show "substitute" behavior—preening, vibrissae-twitching, crouching, etc. In a similar experiment Goldberg (**54**), using an intense flashing light as the conditioned stimulus, observed no tonic-clonic episodes when the light was presented alone. Several running responses were initiated during the presentation of the light but their significance is not clear. Exploration, which had increased, and crouching and preening, which had decreased, during a preliminary phase in which the light was presented without the bell, decreased and increased respectively after the light had been paired with the bell. Frings, *et al.* (**37**) report mice to behave apprehensively, as if "something is going to happen."

Experiments on learned control of the seizure have been more successful. Marx and Van Spankeren (**89**) trained rats to cut off the sound by pushing a pole and then delayed making the pole available in order to determine if control of the stimulus also meant control of the seizure. While there were no significant differences in the number of running attacks and convulsions during the training sequence, the trained had reliably fewer attacks of both sorts during the period of pole-delay testing. During a period of extinction, the trained animals showed a shift back toward their pretraining sensitivity level. Additional indications of control were found in the forepaw clonus that emerged during training and in the attacks that sometimes occurred after the animal had executed the pole response. The experimenters suggest that the animal subjects develop a technique that releases them from a noxious situation and with it a "sense of control" as indicated by a general resumption of eating behavior. In a further study, Goodson and Marx (**55**) substituted wheel-turning for pole-tilting. Members of a control-experimental pair were testing simultaneously in adjacent test chambers. Both were subjected to auditory stimulation and electric shock through the feet. The experimental animal, however, had available in its compartment a wheel which, when turned, reduced the strength of the shock current. Comparison of the two groups of *S*s on a seizure-resistant index indicated the experimental animals to be reliably more resistant. When the control animals were given wheel training and the experimentals denied access to the wheel, the resistance relationship was reversed. These results are in line with Griffiths' previously described conclusions on manipulation.

Chance (**89**) meanwhile reports

that the introduction of a plastic hut for a brief period at the beginning of each test provides for immunity. Three of his mice showed searching movements when the hut was not present, creeping to the area where it had been, then sitting relaxed through the sound test. In order to check on the possibility that the hut reduced stimulus intensity, a plastic horseshoe was substituted for the hut with no change in results. This learned refractoriness is regarded as responsible for the wide intra- and intersubject variability in susceptibility. It most readily supports the general class of psychological seizure theories.

AUDIOGENIC CONVULSIONS AS AN INDEPENDENT VARIABLE

Some Effects upon Bodily Processes

A few papers describe more lasting physiological changes than those previously associated with the period of testing and/or convulsion. Patton, Russell, and Pierce (**117**) report increased sensitivity to electroshock convulsions following a series of daily exposures to subconvulsive auditory stimulation. Yeakel, *et al.* (**143**) exposed rats to five minutes of subconvulsive air blasts daily for one year and at random intervals thereafter and report persistent hypertension in the subjects during the last two to four months of treatment. During the first two months the mean systolic pressure in their controls was 113 in contrast to the experimentals' 124 mm. Hg., during the last two months 127 in contrast to 154. While no morphological change was apparent in the adrenal glands of the blasted animals, their average gland weight tended to be greater. In a second experiment (**79**) adrenalectomy was followed by slightly subnormal blood pressure (97 mm. Hg.), replacement therapy restoring the normal level.

To investigate possible seizure-connected pituitary dysfunction, Nolan (**107**) examined virgin female rats for signs of pseudopregnancy during a series of 21 audiogenic seizure tests. No change in length of oestrus cycle was observed in the convulsing experimentals, nonconvulsing experimentals, or unexposed controls. Exposure to electroshock convulsion also failed to induce the condition, a result at variance with the previous report of Jensen and Stainbrook (**73**). Ginsburg and Huth (**49**) have found less free acetylcholine in the submaxillary glands of DBA mice than in C57's. No differences, however, were found in the acetylcholine level in whole brain extracts of these strains.

Death during seizures appears as a risk in rats only if they have been subjected to severe nutritional deficiency or are very young. That it is, in contrast, a common outcome with certain strains of mice has already been discussed.

Postseizure Changes in Behavior

Reports of changes in emotional tone subsequent to a series of seizures continue to appear in the literature (**8, 54**). Rats initially gentle and curious, after only one or two exposures, begin to display premonitory symptoms when placed in the test chamber. Exploratory behavior ceases, the animal crouches, shivers, and displays weaving head movements. He may become more inactive and "retiring" but more irritable and jumpy and resist being removed from his home-cage and/or placed in the test situation. Similar changes have been described for mice (**37**). In addition, nonconvulsing rats are characterized as more compulsive (**82**). James and Boyles (**72**) found a significant reduction in food and water intake in rats following convulsion, normal levels being restored within 58 hours. This they attributed to the temporary drop in activity that generally follows convulsions.

Maier (**82**) claims that convulsing rats breed less readily, a condition that might be expected to accompany a state of chronic anxiety. Still there are no more recent data to contradict the findings of Farris and Yeakel (**25**, **26**). Frings, Frings, and Kivert (**37**) compared litter size and survival in three mouse strains: mongrel Swiss Albinos produced large litters and reared most to testing age, C57's had smaller litters but reared most, DBA's had the fewest offspring per litter and reared only two thirds of these.

Hall and Whiteman (**66**) exposed DBA×C57 hybrids to four sound tests at the age of four to seven days, then obtained measures of emotionality with the open-field test at 30–40 days, the stovepipe test at 70–80 days, and the open-field test again at 100–110 days. The experimentals were judged more emotional than controls not exposed to infantile stimulation, but no reliable difference was apparent after the first test. This was attributed to familiarization with emotionality-test situations. Challenging these results Griffiths and Stringer (**63**) exposed rats to auditory stimulation, rapid rotation, shifts in temperature, and electric shock until 57 days of age. At 60 days observations were begun of behavior in the open-field situation, susceptibility to audiogenic attack, and performance in an 11-blind Warner-Warden maze and Lashley jumping apparatus. At no time were there significant differences between the experimental animals and unexposed controls on the tests.

Chance (**17**) calls attention to what he refers to as a new component of the postconvulsive behavior of Peromyscus. While testing mice in pairs, he observed that the first animal to recover from the epileptoid phase frequently attacked its test partner, holding with its forepaws and vigorously biting. Such attacks, however, were also directed toward a stuffed mouse pelt, a cotton-wool roll, and a plasticine model, the roll and the fur being attacked preferentially. The response is well coordinated and hence regarded as an automatism. The present writer has occasionally observed similar behavior in rats when attacks were followed by hyperexcitement.

A small number of papers deal with possible effects of exposure to intense sound and/or convulsions upon performance in routine learning situations. While these serve to clarify certain previously reported relationships they add little that is new. Brady, Stebbins, and Galambos (**10**) trained rats to make a lever-pressing response and then conditioned an emotional response (suppression of the lever response by a combination of click and shock). After the sixth conditioning trial, a sequence of auditory tests, 21 on alternate days and 15 daily, was begun. After a three-day interval retention of the conditioned emotional reaction was checked by giving one nonreinforced trial. This last was almost completely abolished in animals with 30 or more seizures, markedly reduced in animals giving from 1 to 26 convulsive reactions, but suffered no change in seizure-immune animals. The rank-order correlation between number of convulsions and an index reflecting strength of lever pressing response was .89, indicating the importance of convulsions, rather than exposure to sound, for performance.

Different conclusions are suggested by Shaw, Utecht, and Fullanger's (**126**) experiment. While both studies demonstrate a clear difference between convulsives, on the one hand, and nonconvulsive and nonexposed controls, on the other, the latter experiment does not indicate defective retention to be responsible for this. Thirty-seven rats (16 controls, 12 nonconvulsing and 9 convulsing experimentals) were trained in a Tolman, Ritchie, and Kalish place-response maze, modified by attaching tunnels to the alleys near uncovered food boxes. The experimentals were then exposed to 5 minutes of key jingling. After 25 minutes three retest trials were run at 2-minute intervals. While there were no differences among the three groups of *S*s on initial learning, the convulsers took markedly longer on the retest trials. There were, however, no between-groups differences in the mean numbers of errors during the retest sequence. None of the animals hesitated at the choice-point. The convulsives, meanwhile, "hid" in the tunnel, only two of the nine going on to the food box on any of the 10-minute retest trials. This failure to go to food after correct choice is explained as the superposition of a fear response upon a learned pattern and is hard to reconcile with the Brady interpretation of poor retention due to the disorganization of new traces resulting from violent convulsive activity.

Similar to the Shaw, Utecht, and Fullanger results are those of Stanibrook (**129**) who reported that 30 three-minute sound tests, yielding an average of 62 per cent convulsions, were followed by prolonged times on at least five trials of relearning a maze. There was, however, no increment in errors. The behavior of the rats on the first several relearning trials was characterized as hypervigilant and sensitive to fright. In a further experiment utilizing a water maze, "noise-fright" rats most frequently displayed VTE at the last learned choice-point, suggesting that the disrupting effect of fright is greatest upon more recently acquired reactions. Thirty days after sound tests were terminated, no detrimental influences were observed in maze performance. In line with this and the results of Griffiths and Stringer (**63**) previously described, Gilbert and Gawain (**46**) and Martin and Hale (**105**) find no clear evidence of decrement in Warner-Warden maze performance for nonseizure rats exposed to sound varying from 5.5 to 33 kc in frequency, 110 to 147 db in intensity, and up to 15 minutes in duration.

Shohl (**127**) reports no effect of exposure to sound per se upon acquisition of a brightness-discrimination habit when comparison involved conventional time and error criteria. However, qualitative differences were apparent in the behavior of certain animals on exposure days: they displayed jerky running, exaggerated hesitation, and posturing. There were, in addition, increased errors, longer starting times and marked position habits on the days during which the subject had convulsed. Generally, position habits appeared only after the animal had had several convulsions. Since poorer performance occurred only on convulsion trials, Shohl suggests the disturbance to be due not to learning, but to either severe emotional upset or temporary metabolic defect. In summary, these few studies prompt the conclusion that defective performance, when it is associated with convulsiveness, derives from a generalized but transient noncognitive change, not from a faulty capacity for learning or retention.

TRENDS IN INTERPRETATION

Not all animals in an experimental population succumb to convulsion-provoking auditory stimulation. Nor do sensitive individuals respond on all occasions. The appearance of a convulsive response, and the form it takes, is contingent upon a variety of factors: the nature and duration of the auditory agent, the application of drugs, the degree of restraint placed upon the animals' movements and/or the extent of manipulatory activity, the animal *S*'s strain, its nutritional status, its body temperature, lesions in its nervous system, what it has been trained to do in the test situation, the presence of middle-ear infection, and possibly its sex, hormone balance, and temperament.

Explanations of seizures have been delineated within both psychological and physiological frameworks. Physiological theories, in general more readily accepted, have derived from observed changes in sensitivity associated with characteristics of the auditory stimulus, binding or otherwise restraining the animal's movements, variations in dietary intake, the administration of drugs, and surgical damage to the nervous system. They are most appropriate for the data on strain and age differences in susceptibility, interanimal similarity in seizure pattern, and the tendency for responses to occur within about 60 seconds. In their present form they cannot effectively deal with increased susceptibility following conflict or control through learning.

The oldest of the physiological theories (**100**) maintains that the audiogenic seizure is an overflow phenomenon: severe auditory bombardment effects a high level of excitation in auditory centers with resultant spread into adjacent motor centers and widespread reflexive reaction. Consonant with this interpretation is the superiority of noise over pure tones, high auditory frequency and intensity over low frequency and intensity, the data on priming, and the presence, almost without exception, of sound with the occurrence of convulsions. The effectiveness of binding in reducing seizure is attributed to straining against the bonds with the consequent reduction of motor tension before it reaches a convulsive level. Inconsistent with this is the immunity of completely immobile animals and the inevitable running that presages the tonic-clonic attack.

Though intense sound be assumed to be a necessary condition to seizure—and this is yet to be unequivocally established—the overflow model need not be the simple one just described. A more elaborate model meanwhile requires more precise specification of the role of auditory excitation. Several possibilities present themselves: critical level overflow may originate in more than one sensory center; auditory input may serve to detonate intraorganismic changes that in turn lead to overflow. Experimental check on the vestibule as a source of positive feedback has yielded negative results (**28**), although data on the latency of convulsion and the duration of running in animals with dorsal hemisections suggest kinesthesis as a contributory factor (**8**). The work on middle-ear infection points to pain as an important modality, though direct experimental evidence is still lacking (**88**). Meanwhile, in recent years, the importance of the autonomic system and subcortical centers, has become increasingly clear. The few data available on cerebral dehydration, damage to the frontal areas, motor cortex and pyramidal tracts, sectioning of the ninth, tenth, and eleventh cranial nerves prompt extension of the overflow model in terms of the extrapyramidal system.

Audiogenic seizures may also be characterized as a depletion phenomenon, for they are precipitated most readily in animals suffering from dietary restriction or reduced oxygen tension. There is, however, no fundamental opposition between this and the overflow hypothesis, for overflow is a function not only of magnitude of stimulation but also of sensitivity of the nervous system and any condition or conditions that result in enhanced sensitivity make for overflow, or faulty operation of a governor that regulates overflow. The existence of genetic differences and the data on priming suggest that the defect may be of a rather specific nature such as is found in the relationship between phenylpyruvic acid metabolism and oligophrenia. Already a detailed model has been projected dealing with the protective effects of glutamic acid supplements.

Two bits of data warrant further exploitation: the claim that death in DBA mice results from respiratory failure, and reduced clotting times in susceptible animals. It may be highly important to determine if this latter effect is peculiar to the blood or occurs also in other body fluids. Should clots occur during the period of vigorous running, they could conceivably render the cerebrum anoxic and release the brain stem from cortical control with resultant convulsion. This speculation requires assuming that auditory input serves a detonator function. This has yet to be investigated.

Despite general preference for a physiological model, nothing has been done to assay systematically physiological differences among susceptible and nonsusceptible animals. Although electroshock convulsions have been discussed within the framework of Selye's adaptation syndrome hypothesis (**122**), and although audiogenic seizures have been regarded as a stress phenomenon by both physiologically and psychologically biased investigators, no vigorous attempt to exploit Selye's logic has been made for the latter phenomenon. Meanwhile, Bevan (**6**) after repeatedly placing animals in laboratory situations which he characterizes as "conflict" and "fear provoking," reports changes in hemoglobin level, platelet and erythrocyte count, spleen size, and the presence of mucus and blood in feces and urine. The success of Williams and his colleagues (**140**) in defining "metabolic personality" in alcoholism, schizophrenia, and feeblemindedness recommends his experimental approach to the student of audiogenic seizures. Another lead that might be profitably followed involves investigating the effect of systematically varying the chemical properties of convulsants and anticonvulsants.

The original theory of audiogenic seizures was a conflict theory derived to account for the convulsions observed in some rat *S*s when forced with an air blast to respond in the Lashley jumping situation (**81**). According to Maier (**81**) the animal, trapped between two negative incentives, is rendered immobile by the simultaneous arousal of incompatible response tendencies. This results in the accumulation of tension, which if it reaches "catastrophic" level, induces convulsion. This is an overflow model elaborated in both physiological and psychological terms. The noise-fright, pain, emotionality, and escape hypotheses may all be interpreted as versions of it. It can deal with the enhanced susceptibility in sound tests that follows forced response in the one-card jumping situation; the protective effects of "directed" activity (substitute behavior, conditioned wheel-turning, manipulation of binding or harness, escape into plasticine horseshoes, etc.); apprehensiveness in the test situation; and increase in frequency and change in pattern with repeated tests. It explains why sound is peculiarly effective (although it might be pointed out that this of all modalities requires the least energy for arousal). While it can account for the facilitative effects of priming, it cannot explain the protective effects nor are certain of the more recent data on reduction of movement consonant with it. As was stated earlier, it cannot deal directly with the possible effects of middle-ear infection, damage to the central nervous system (although it is not inconsistent with data suggesting the importance of the extrapyramidal system), and the clear-cut age and strain differences in sensitivity.

A purely psychological conflict model may be substituted for the psychophysiological model just summarized. The convulsive response is then viewed as a positive solution to the problem of escaping a distressful situation (i.e., as an hysterical convulsion) and thus a form of *aus dem Feld gehen* behavior. This interpretation might account for reduced seizures following substitute behavior, running to the locus of the hut, wheel-turning, manipulation of binding and total immobility, the controlled phase in the running sequence, the peculiar effectiveness of auditory stimulation, and the decrease in latency with continuous tests. It would suffer, however, the same inadequacies as the conflict-overflow hypothesis.

Controversy at this time constitutes indulgence in "metascience," for it requires almost complete dependence upon emotional bias for resolution. This is so for a number of reasons. Any inference of functional principles requires a knowledge of independent variables, defined as dimensions. Our present information about audiogenic seizures is, for the most part, simply a listing of conditions which have been found to be associated with susceptibility. Among these data there is much that is contradictory and ambiguous. Furthermore, the indices used to describe susceptibility are rather crude and little has been done to devise more refined measures.

There also appears to be a prevalent assumption that alternative models must be in opposition; if a physiological explanation is adequate for certain data, the possible appropriateness of a psychological theory is excluded. Such an assumption is not valid, for the adequacy of a model is defined not by the number of alternative explanations it eliminates, but by its versatility in integrating data and its fruitfulness for prediction. That models may complement each other is suggested on the one hand by the behavioral data that yields the label, noise-fright reaction, and, on the other, by physiological data that indicates the importance of the extrapyramidal system.

However, in the instances where both extant physiological and psychological models seem appropriate, it is not possible to achieve unambiguous definition of the key concepts. While sound has been a component of every situation in which audiogenic seizures have occurred, it is not possible to describe a sound test which might not also be characterized as a conflict situation.

If available models lack definitiveness, they also are probably overly simple and over-extended. Although it may be established that auditory stimulation is prerequisite to convulsions, it does not necessarily follow that auditory overflow is the crucial condition. Nor because conflict may be construed to be associated with premonitory behavior or the initiation of the running-phase does it mean that it explains the occurrence of the fit. We have yet to consider seriously the possibility of different mechanisms directing different phases of the response—the recent work on convulsive risk, death risk, and priming are exceptions. Perhaps a valuable dictum might be derived by modifying Beach's (3) principles of multiple causation and multiple effects: in no instance is a response due to one and only one factor; nor does any single process or condition have one and only one behavioral result.

Audiogenic seizures constitute an important research problem. For the psychologist they provide a dramatic behavior phenomenon for study, for the neurophysiologist, a situation that may lead to important insights concerning central nervous system function, and for the biochemist, a potentially valuable assay technique—for all, an opportunity for interdisciplinary fertilization.

REFERENCES

1. Abood, L. G., & Gerard, R. W. Phosphorus metabolism and function in the nervous system. *Fed. Proc.*, 1952, **11**, 3.
2. Adams, J. S., & Griffiths, W. J., Jr. The effects of tridione on audiogenic fits in albino rats. *J. comp. physiol. Psychol.*, 1948, **41**, 319–326.
3. Beach, F. A. Body chemistry and perception. In R. R. Blake and G. V. Ramsey, *Perception, an approach to personality*. New York: Ronald, 1951, Ch. 3.
4. Beach, F. A., & Weaver, Thelma. Noise-induced seizures in the rat and their modification by cerebral injury. *J. comp. Neurol.*, 1943, **79**, 379–392.
5. Benedict, Lois L. Sound-induced seizures in rats reared on diets with high sugar content. Unpublished M.S. thesis, Univer. of New Mexico, 1951.
6. Bevan, J. M. Effects of stress upon certain physiological mechanisms and behavior of the albino rat. Unpublished Ph.D. thesis, Duke Univer., 1951.
7. Bevan, W., Jr., Hard, J. S., & Seal, U. S. Sound-induced seizures in rats fed an amino-acid deficient diet. *J. comp. physiol. Psychol.*, 1951, **44**, 327–333.
8. Bevan, W., Jr., & Hunt, E. L. Proprioceptive inflow and susceptibility to experimentally-induced seizures. *J. comp. physiol. Psychol.*, 1953, **46**, 218–224.
9. Bingham, W. E., & Griffiths, W. J., Jr. The effect of different environments during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 307–312.
10. Brady, J. V., Stebbins, W. C., & Galambos, R. The effect of audiogenic convulsions on a conditioned emotional response. *J. comp. physiol. Psychol.*, 1953, **46**, 363–367.
11. Brdar, P., & Moyer, P. The relationship of the glossopharyngeal and vagus nerves to audiogenic seizures. *Persona*, 1949, **1**, No. 2, 15–17.
12. Busquet, H., & Vischniac, C. Identité d'action du principe sympathomimétique du genet et de l'adrénaline sur la perméabilité capillaire et sur la crise audiogène. *C.R. Soc. Biol., Paris*, 1950, **144**, 398–400.
13. Cain, J., & Mercier, J. Influence de certain médicaments anticonvulsivants sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1948, **142**, 688–691.
14. Cain, J., Mercier, J., & Corriol, J. Sur electroencephalogramme du rat albinos soumis à la crise audiogène. *C. R. Soc. Biol., Paris*, 1951, **148**, 915–916.
15. Cain, J., & Naquet, R. Influence de la contention et des circonstances expérimentales sur la crise audiogene du rat. *C.R. Soc. Biol., Paris*, 1950, **144**, 1192–1193.
16. Chance, M. R. A. Factors influencing the toxicity of sympathomimetic amines to solitary mice. *J. pharm. exp. Therapeut.*, 1947, **89**, 289–296.
17. Chance, M. R. A. Aggression, a com-

ponent of post-epileptic automatism in Peromyscus. *Nature*, 1948, **161**, 101–102.

18. Chance, M. R. A. The suppression of audiogenic hyper-excitement by learning in Peromyscus maniculatus. Unpublished manuscript, Univer. of Birmingham, England.
19. Chance, M. R. A., & Yaxley, D. C. New aspects of the behavior of Peromyscus under audiogenic hyper-excitement. *Behavior*, 1949, **2**, 96–105.
20. Chance, M. R. A., & Yaxley, D. C. Central nervous function and changes in brain metabolite concentration. I. Glycogen and lactate in convulsing mice. *J. exp. Biol.*, 1950, **27**, 311–323.
21. Colman, K. W. Conditioning and audiogenic seizure. *Psi Chi Newsletter*, Oct. 1952, 13–14.
22. Deese, J., & Morgan, C. T. Comparative and physiological psychology. In C. P. Stone, and D. W. Taylor (Eds.), *Ann. Rev. Psychol.*, **2**, 1951, 205–208.
23. Dember, W. N., Ellen, P., & Kristofferson, A. B. The effects of alcohol on seizure behavior in rats. *Quart. J. Stud. Alcohol.*, 1953, **14**, 390–394.
24. Dice, L. R., & Barto, Elizabeth. Ability of mice of the genus Peromyscus to hear ultrasonic sounds. *Science*, 1952, **116**, 110–111.
25. Farris, E. J., & Yaekel, Eleanor H. Effects of a neurophysiological stimulus on the breeding of albino rats. *Anat. Rec.*, 1942, **84**, 454.
26. Farris, E. J., & Yaekel, Eleanor H. Breeding and rearing of young by albino rats subjected to auditory stimulation. *Anat. Rec.*, 1944, **89**, 325-330.
27. Finger, F. W. Convulsive behavior in the rat. *Psychol. Bull.*, 1947, **44**, 201–248.
28. Finger, F. W., Bice, R. C., & Day, W. F. Audiogenic seizures and streptomycin-induced vestibular dysfunction. *J. comp. physiol. Psychol.*, 1952, **45**, 163–169.
29. Frings, H., & Frings, Mabel. Otitis media and audiogenic seizures in mice. *Science*, 1951, **113**, 689–690.
30. Frings, H., & Frings, Mabel. The development of strains of albino mice with predictable susceptibility to audiogenic seizures. *Anat. Rec.*, 1952, **113**, 550–551.
31. Frings, H., & Frings, Mabel. Acoustical determinants of audiogenic seizure in laboratory mice. *J. acoust. soc. Amer.*, 1952, **24**, 163–169.
32. Frings, H., & Frings, Mabel. Audiogenic seizures in the laboratory mouse. *J. Mammal.*, 1952, **33**, 80–87.
33. Frings, H., & Frings, Mabel. Latent periods of audiogenic seizure in mice. *J. Mammal.*, 1952, **33**, 487–491.
34. Frings, H., & Frings, Mabel, Development of strains of albino mice with predictable susceptibilities to audiogenic seizures. *Science*, 1953, **117**, 283–284.
35. Frings, H., & Frings, Mabel. The production of stocks of albino mice with predictable susceptibility to audiogenic seizure. *Behaviour*, 1953, **5**, 305–319.
36. Frings, H., Frings, Mabel, Fuller, J. L., Ginsburg, D. E., Ross, S., and Vicari, E. M. Standardization of nomenclature describing audiogenic seizures in mice. *Behaviour*, 1952, **4**, 157–160.
37. Frings, H., Frings, Mabel, & Kivert, A. Behavior patterns of the laboratory mouse under auditory stress. *J. Mammal.*, 1951, **32**, 60–76.
38. Frings, H., & Kivert, A. Nicotine facilitation of audiogenic seizures in laboratory mice. *J. Mammal.*, 1953, **34**, 391–392.
39. Fuller, J. L. Genetic controls of audiogenic seizures in hybrids between DBA subline 2 and C57 Bl. subline 6. *Rec. genet. soc. Amer.*, 1949, **18**, 86–87.
40. Fuller, J. L., Easler, Clarice, & Smith, Mary E. Inheritance of audiogenic seizure susceptibility in the mouse. *Genetics*, 1950, **35**, 622–632.
41. Fuller, J. L., & Ginsburg, B. E. Effect of adrenalectomy on the anticonvulsant action of glutamic acid in mice. *Amer. J. Physiol.*, 1954, **173**, 367–370.
42. Fuller, J. L., & Rappaport, Adrienne. The effect of wetting on sound-induced convulsions in mice. *J. comp. physiol. Psychol.*, 1952, **45**, 246–249.
43. Fuller, J. L., & Smith, Mary E. Kinetics of the mass action of the mammalian nervous system by sound. *Anat. Rec.*, 1952, **113**, 547.
44. Fuller, J. L., & Smith, Mary E. Kinetics of sound-induced convulsions in some inbred mouse strains. *Amer. J. Physiol.*, 1953, **172**, 661–670.
45. Fuller, J. L., & Williams, Elizabeth. Gene-controlled time constants in convulsive behavior. *Proc. nat. Acad. Sci., Wash.*, 1951, **37**, 349–356.
46. Gilbert, P. F., & Gawain, G. C. V.

Sonic and ultrasonic effects on maze learning and retention in the albino rat. Unpublished Ph.D. thesis, Pennsylvania State Coll., 1950.

47. GINSBURG, B. E., & FULLER, J. L. A comparison of chemical and mechanical alterations of seizure patterns in mice. *J. comp. physiol. Psychol.*, 1954, **47**, 344–348.

48. GINSBURG, B. E., & HOVDA, R. B. On the physiology of gene-controlled audiogenic seizures in mice. *Anat. Rec.*, 1947, 99, 621.

49. GINSBURG, B. E., & HUTH, E. Some aspects of the physiology of gene-controlled audiogenic seizures in inbred strains of mice. *Genetics*, 1947, **32**, 87.

50. GINSBURG, B. E., MILLER, DOROTHY, & ZAMIS, MILDRED J. On the mode of inheritance of susceptibility to sound-induced seizures in the house mouse (*Mus musculus*). *Genetics*, 1950, **35**, 109.

51. GINSBURG, B. E., and ROBERTS, E. Glutamic acid and central nervous system activity. *Anat. Rec.*, 1951, **111**, 492.

52. GINSBURG, B. E., ROSS, S., ZAMIS, MILDRED J., and PERKINS, AGNES. An assay method for the behavioral effects of 1-glutamic acid. *Science*, 1950, **112**, 12–13.

53. GINSBURG, B. E., ROSS, S., ZAMIS, MILDRED J., and PERKINS, AGNES. Some effects of 1+glutamic acid on sound-induced seizures in mice. *J. comp. physiol. Psychol.* 1951, **44**, 134–141.

54. GOLDBERG, DOROTHY. Audiogenic seizures and related behavior in the albino rat. Unpublished M.A. thesis, Emory Univer., 1952.

55. GOODSON, F. E., & MARX, M. H. Increased resistance to audiogenic seizure in rats trained on an instrumental wheel-turning response. *J. comp. physiol. Psychol.*, 1953, **46**, 225–230.

56. GREENBERG, L. A., & LESTER, D. The effect of alcohol on audiogenic seizures in rats. *Quart. J. Stud. Alcohol.*, 1953, **14**, 385–389.

57. GRIFFITHS, W. J., JR. Audiogenic fits produced by magnesium deficiency in tame domestic Norway rats and in wild Norway and Alexandrine rats. *Amer. J. Physiol.*, 1947, **149**, 135–141.

58. GRIFFTHS, W. J., JR. The effects of adrenalectomy on the incidence of audiogenic seizures in rats. *Amer. Psychologist*, 1948, **3**, 332–333. (Abstract)

59. GRIFFITHS, W. J., JR. Effect of adrenalectomy on incidence of audiogenic seizures among domestic and wild rats. *J. comp. physiol. Psychol.*, 1949, **42**, 303–312.

60. GRIFFITHS, W. J., JR. Self-selection of diet in relation to audiogenic seizures in rats. *Science*, 1950, **112**, 786.

61. GRIFFITHS, W. J., JR. The influence of behavioral factors in the incidence of audiogenic seizures in rats. *J. comp. physiol. Psychol.*, 1953, **46**, 150–152.

62. GRIFFITHS, W. J., JR., & COHEN, M. Effects of morphine sulphate and diaminon hydrochloride on incidence of audiogenic seizures in albino rats. *J. comp. physiol. Psychol.*, 1949, **42**, 391–397.

63. GRIFFITHS, W. J., JR., & STRINGER, W. F. The effects of intense stimulation experienced during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 301–306.

64. HALL, C. S. Genetic differences in audiogenic seizures. *J. Hered.*, 1947, **38**, 3–6.

65. HALL, C. S. The genetics of behavior. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 309–310, 317–319.

66. HALL, C. S., & WHITEMAN, P. H. The effects of infantile stimulation upon later emotional stability in the mouse. *J. comp. physiol. Psychol.*, 1951, **44**, 61–66.

67. HAMILTON, H. C., & MAHER, EILEEN B. The effects of glutamic acid on the behavior of the white rat. *J. comp. physiol. Psychol.*, 1947, **40**, 463–468.

68. HARLOW, H. F. Physiological psychology. In V. E. Hall, (Ed.), *Ann. Rev. Physiol.*, 1949, 284–286.

69. HOFELD, J. R. Changes in metabolic rate during audiogenic seizure in the rat. *Amer. Psychologist*, 1948, **4**, 359. (Abstract)

70. HURDER, W. P., & SANDERS, A. F. Audiogenic seizure and the adrenal cortex. *Science*, 1953, **117**, 324–326.

71. HURDER, W. P., & SANDERS, A. F. A correction and additional observation. *Science*, 1954, **119**, 476–477.

72. JAMES, W. T., & BOYLES, C. The effect of seizure in rats on food-and-water intake. *Amer. J. Psychol.*, 1950, **63**, 284–286.

73. JENSEN, G. D., & STAINBROOK, E. The effects of electrogenic convulsions on the estrus cycle and weight of rats. *J. comp. physiol. Psychol.*, 1949, 42, 502–505.

74. Kenshalo, D. R., & Kryter, K. D. Middle ear infection and sound-induced seizures in rats. *J. comp. physiol. Psychol.*, 1949, **42**, 328–331.
75. Lazovik, A. D., and Patton, R. A. The relative effectiveness of auditory stimulation and motivational stress in precipitating convulsions associated with magnesium deficiency. *J. comp. physiol. Psychol.*, 1947, **40**, 191–202.
76. Lindsley, D. B., Finger, F. W., & Henry, C. W. Some phsyiological aspects of audiogenic seizures in rats. *J. Neurophysiol.*, 1942, **5**, 185–198.
77. Lindzey, G. Emotionality and audiogenic seizure susceptibility in five inbred strains of mice. *J. comp. physiol. Psychol.*, 1951, **44**, 389–393.
78. Longhurst, Joan U. Effects of brain injury to the rat on seizures produced during auditory stimulation. Unpublished Ph.D. thesis, Univer. of Michigan, 1948.
79. McCann, S. M., Rothballer, H. B., Yeakel, Eleanor H., & Shenkin, H. A. Adrenalectomy and blood pressure of rats subjected to auditory stimulation. *Amer. J. Physiol.*, 1948, **155**, 128–131.
80. McGrael, L. Pure tone and audiogenic seizures in mice. Unpublished M.S. thesis, Pennsylvania State Coll., 1949.
81. Maier, N. R. F. *Studies of abnormal behavior in the rat.* New York: Harper, 1939.
82. Maier, N. R. F. Experimentally induced abnormal behavior. *Sci. Mon.*, 1948, **67**, 211–216.
83. Maier, N. R. F. *Frustration, the study of behavior without a goal.* New York: McGraw-Hill, 1949, 134–140.
84. Maier, N. R. F., Feldman, R. S., & Longhurst, Joan U. Studies of abnormal behavior in the rat. XX. Change in seizure patterns with repeated testing. *J. comp. physiol. Psychol.*, 1947, **40**, 73–86.
85. Maier, N. R. F., & Longhurst, Joan U. Studies of abnormal behavior in the rat. XXI. Conflict and "audiogenic" seizure. *J. comp. physiol. Psychol.*, 1947, **40**, 397–412.
86. Maier, N. R. F., Longhurst, Joan U., & Ellen, P. Effects of lactose in the diet on seizure behavior of male and female rats. *J. comp. physiol. Psychol.*, 1951, **44**, 501–506.
87. Marx, M. H., & Chambers, R. M. Incidence of audiogenic seizure following experimental induction of middle ear disorder. *J. comp. physiol. Psychol.*, 1952, **45**, 239–245.
88. Marx, M. H., & Jurko, Marion. The relationship between the audiogenic seizure and middle ear disease. *J. genet. Psychol.*, 1950, **76**, 221–239.
89. Marx, M. H., & Van Spankeren, W. J. Control of the audiogenic seizure by the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 170–180.
90. Mercier, J. Note préliminaire sur l'influence exercée par quelques medicaments sympathicométique sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1949, **143**, 1125–1127.
91. Mercier, J. Structure moléculaire et pouvoir anticonvulsivant. Influence exercée par le gardénal, le metabromogardénal, et le bromure de sodium sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1950, **144**, 1174–1176.
92. Mercier, J. Structure moléculaire et pouvoir anticonvulsivant. Influence des substitutions dans les noyaux, de l'acide barbiturique et de l'hydanotoine. *C.R. Soc. Biol., Paris*, 1950, **144**, 1177–1678.
93. Mercier, J. Influence exercée par deux sympathicolytiques sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1951, **145**, 906–908.
94. Mercier, J., & Cain, J. Influence exercée par la tridione (trimethyloxazolidinedione) sur les crises audiogènes du rat albinos. *C.R. Soc. Biol., Paris*, 1948, **142**, 993–994.
95. Mercier, J., & Garnier, J. Influence exercée par l'hydratation du tissue cérébral sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris.* 1951, **145**, 1199–1201.
96. Michels, K. M., & Bevan, W. Audiogenic seizures in rats as a function of the volume of the test chamber. *J. genet. Pyschol.*, 1952, **81**, 185–192.
97. Miller, Dorothy S., & Zamis, Mildred J. Independent occurrences of audiogenic seizures and middle-ear infection in inbred mice. *Anat. Rec.*, 1949, **105**, 556–557.
98. Miller, Dorothy S., Ginsburg, B. E., & Potas, Mildred Z. Inheritance of seizure susceptibility in the house mouse (*Mus musculus*). *Genetics*, 1952, **37**, 605–606.
99. Morgan, C. T., & Stellar, E. Physiological Psychology. New York: McGraw-Hill, 1950, 523–527.
100. Morgan, C. T., & Waldman, H. "Con-

flict" and audiogenic seizures. *J. comp. Psychol.*, 1941, **31**, 1–11.

101. Morin, G., & Cain, J. Sur un comportement anormal provoqué par certains bruits chez le rat. *C.R. Soc. Biol., Paris*, 1947, **141**, 1245–1247.
102. Morin, G., & Cain, J. Sur la signication des convulsions et des attitudes catatonique provoquées par le son chez rat. *C.R. Biol., Paris*, 1947, **141**, 1247–1248.
103. Morin, G., & Cain, J. Une forme peu connue d'epilepsie experimentale: l'-épilepsie audiogène. *La Medicine*, 1949, **30**, 12–16.
104. Morin, G., Canac, F., Cain, J., & Gavreau, V. Obtention des crises audiogènes du rat par des sons definis. *C.R. Soc. Biol., Paris*, 1948, **142**, 359–360.
105. Morton, J., & Hale, E. B. The effect of high intensity sound on maze learning in the albino rat. Report submitted to Bio-Acoustic Unit, Aeromedical Lab., Washington, D.C.: Wright-Patterson A.F.B., Ohio, 1953.
106. Munn, N. C. *Handbook of psychological research on the rat.* Boston: Houghton Mifflin, 1950, 420–439.
107. Nolan, C. Y. Stimulational seizures without pseudopregnancy in white rats. *J. comp. physiol. Psychol.*, 1952, **45**, 183–187.
108. Norton, H. W. On audiogenic seizure and the adrenal cortex. *Science*, 1954, **119**, 475–476.
109. Patton, R. A. Nutritional deficiency and sound-induced convulsive seizures in the rat. *Ment. Health Bull.*, 1946, **24**, 16–20.
110. Patton, R. A. The incidence of middle-ear infection in rats susceptible to sound induced seizures. *Amer. Psychologist*, 1947, **2**, 320. (Abstract)
111. Patton, R. A. Purulent otitis media in albino rats susceptible to sound-induced seizures. *J. Psychol.*, 1947, **24**, 313–317.
112. Patton, R. A. Sound-induced convulsions in the hamster associated with magnesium deficiency. *J. comp. physiol. Psychol.*, 1947, **40**, 283–289.
113. Patton, R. A. Vitamin B-complex concentrates and the incidence of sound-induced seizures in young albino rats maintained on purified diets. *J. comp. physiol. Psychol.*, 1947, **40**, 323–332.
114. Patton, R. A. The experimental approach to convulsive seizures. *Ment. Health, Bull.*, 1947, **25**, 15.
115. Patton, R. A. Maternal nutritional deficiency and the incidence of sound-induced convulsions in young albino rats. *Proc. Amer. Psychopath. Assn.*, 1947, **36**, 204–209.
116. Patton, R. A. Abnormal behavior in animals. In C. P. Stone (Ed.), *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951, 479–498.
117. Patton, R. A., Russell, R. W., & Pierce, J. F. The effects of auditory stimulation on the electro-convulsive threshold. *Amer. Psychologist*, 1949, **4**, 233. (Abstract)
118. Patton, R. A., & Zabarenko, L. M. Further observations on the incidence of middle-ear infection and sensitivity to sound-induced convulsive seizures in young albino rats. Unpublished manuscript, Univer. of Pittsburgh. (Cited in C. P. Stone [Ed.], *Comparative psychology* [3rd Ed.]. New York: Prentice-Hall, 1951, p. 492.)
119. Pilgrim, F. J., & Patton, R. A. Reversibility of sound-induced convulsions associated with a pyridoxin deficiency. *Amer. Psychologist*, 1948, **3**, 359. (Abstract)
120. Pilgrim, F. J., & Patton, R. A. Production and reversal of sensitivity to sound-induced convulsions associated with a pyridoxin deficiency. *J. comp. physiol. Psychol.*, 1949, **42**, 422–426.
121. Rabe, P. L. The cumulative frustration effect in the audiogenic seizure syndrome of DBA mice. *J. genet. Psychol.*, 1952, **81**, 3–17.
122. Reiss, B. F. The theoretical bases of convulsive therapy in relation to animal experimentation. *J. Pers.*, 1948, **17**, 9–15.
123. Russell, R. W. Comparative psychology. In C. P. Stone (Ed.), *Ann. Rev. Psychol.*, 1954, **5**, 234–237.
124. Scott, J. P. The use as test material of inbred strains of mice having high frequencies of audiogenic seizures. *Science*, 1950, **111**, 583.
125. Shafer, J. N., & Meyer, D. R. Effects of intense light stimulation on sound-induced seizures in rats. *J. comp. physiol. Psychol.*, 1953, **46**, 305–306.
126. Shaw, W. A., Utecht, A. J., & Fullanger, E. A. The effect of auditory stimulation upon the immediate retention of a previously learned maze behavior in the albino rat. *J. comp. physiol. Psychol.*, 1953, **46**, 212–215.
127. Shohl, Jane. Effects of exposure to sound on discrimination performance in the rat. *Psychol. Monogr.*, 1951, **65**, No. 12 (Whole No. 329).

74. Kenshalo, D. R., & Kryter, K. D. Middle ear infection and sound-induced seizures in rats. *J. comp. physiol. Psychol.*, 1949, **42**, 328–331.
75. Lazovik, A. D., and Patton, R. A. The relative effectiveness of auditory stimulation and motivational stress in precipitating convulsions associated with magnesium deficiency. *J. comp. physiol. Psychol.*, 1947, **40**, 191–202.
76. Lindsley, D. B., Finger, F. W., & Henry, C. W. Some phsyiological aspects of audiogenic seizures in rats. *J. Neurophysiol.*, 1942, **5**, 185–198.
77. Lindzey, G. Emotionality and audiogenic seizure susceptibility in five inbred strains of mice. *J. comp. physiol. Psychol.*, 1951, **44**, 389–393.
78. Longhurst, Joan U. Effects of brain injury to the rat on seizures produced during auditory stimulation. Unpublished Ph.D. thesis, Univer. of Michigan, 1948.
79. McCann, S. M., Rothballer, H. B., Yeakel, Eleanor H., & Shenkin, H. A. Adrenalectomy and blood pressure of rats subjected to auditory stimulation. *Amer. J. Physiol.*, 1948, **155**, 128–131.
80. McGrael, L. Pure tone and audiogenic seizures in mice. Unpublished M.S. thesis, Pennsylvania State Coll., 1949.
81. Maier, N. R. F. *Studies of abnormal behavior in the rat.* New York: Harper, 1939.
82. Maier, N. R. F. Experimentally induced abnormal behavior. *Sci. Mon.*, 1948, **67**, 211–216.
83. Maier, N. R. F. *Frustration, the study of behavior without a goal.* New York: McGraw-Hill, 1949, 134–140.
84. Maier, N. R. F., Feldman, R. S., & Longhurst, Joan U. Studies of abnormal behavior in the rat. XX. Change in seizure patterns with repeated testing. *J. comp. physiol. Psychol.*, 1947, **40**, 73–86.
85. Maier, N. R. F., & Longhurst, Joan U. Studies of abnormal behavior in the rat. XXI. Conflict and "audiogenic" seizure. *J. comp. physiol. Psychol.*, 1947, **40**, 397–412.
86. Maier, N. R. F., Longhurst, Joan U., & Ellen, P. Effects of lactose in the diet on seizure behavior of male and female rats. *J. comp. physiol. Psychol.*, 1951, **44**, 501–506.
87. Marx, M. H., & Chambers, R. M. Incidence of audiogenic seizure following experimental induction of middle ear disorder. *J. comp. physiol. Psychol.*, 1952, **45**, 239–245.
88. Marx, M. H., & Jurko, Marion. The relationship between the audiogenic seizure and middle ear disease. *J. genet. Psychol.*, 1950, **76**, 221–239.
89. Marx, M. H., & Van Spankeren, W. J. Control of the audiogenic seizure by the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 170–180.
90. Mercier, J. Note préliminaire sur l'influence exercée par quelques medicaments sympathicométique sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1949, **143**, 1125–1127.
91. Mercier, J. Structure moléculaire et pouvoir anticonvulsivant. Influence exercée par le gardénal, le metabromogardénal, et le bromure de sodium sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1950, **144**, 1174–1176.
92. Mercier, J. Structure moléculaire et pouvoir anticonvulsivant. Influence des substitutions dans les noyaux, de l'acide barbiturique et de l'hydanotoine. *C.R. Soc. Biol., Paris*, 1950, **144**, 1177–1678.
93. Mercier, J. Influence exercée par deux sympathicolytiques sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris*, 1951, **145**, 906–908.
94. Mercier, J., & Cain, J. Influence exercée par la tridione (trimethyloxazolidinedione) sur les crises audiogènes du rat albinos. *C.R. Soc. Biol., Paris*, 1948, **142**, 993–994.
95. Mercier, J., & Garnier, J. Influence exercée par l'hydratation du tissue cérébral sur la crise audiogène du rat albinos. *C.R. Soc. Biol., Paris.* 1951, **145**, 1199–1201.
96. Michels, K. M., & Bevan, W. Audiogenic seizures in rats as a function of the volume of the test chamber. *J. genet. Pyschol.*, 1952, **81**, 185–192.
97. Miller, Dorothy S., & Zamis, Mildred J. Independent occurrences of audiogenic seizures and middle-ear infection in inbred mice. *Anat. Rec.*, 1949, **105**, 556–557.
98. Miller, Dorothy S., Ginsburg, B. E., & Potas, Mildred Z. Inheritance of seizure susceptibility in the house mouse (*Mus musculus*). *Genetics*, 1952, **37**, 605–606.
99. Morgan, C. T., & Stellar, E. Physiological Psychology. New York: McGraw-Hill, 1950, 523–527.
100. Morgan, C. T., & Waldman, H. "Con-

flict" and audiogenic seizures. *J. comp. Psychol.*, 1941, **31**, 1–11.

101. Morin, G., & Cain, J. Sur un comportement anormal provoqué par certains bruits chez le rat. *C.R. Soc. Biol., Paris*, 1947, **141**, 1245–1247.
102. Morin, G., & Cain, J. Sur la signication des convulsions et des attitudes catatonique provoquées par le son chez rat. *C.R. Biol., Paris*, 1947, **141**, 1247–1248.
103. Morin, G., & Cain, J. Une forme peu connue d'epilepsie experimentale: l'-épilepsie audiogène. *La Medicine*, 1949, **30**, 12–16.
104. Morin, G., Canac, F., Cain, J., & Gavreau, V. Obtention des crises audiogènes du rat par des sons definis. *C.R. Soc. Biol., Paris*, 1948, **142**, 359–360.
105. Morton, J., & Hale, E. B. The effect of high intensity sound on maze learning in the albino rat. Report submitted to Bio-Acoustic Unit, Aeromedical Lab., Washington, D.C.: Wright-Patterson A.F.B., Ohio, 1953.
106. Munn, N. C. *Handbook of psychological research on the rat.* Boston: Houghton Mifflin, 1950, 420–439.
107. Nolan, C. Y. Stimulational seizures without pseudopregnancy in white rats. *J. comp. physiol. Psychol.*, 1952, **45**, 183–187.
108. Norton, H. W. On audiogenic seizure and the adrenal cortex. *Science*, 1954, **119**, 475–476.
109. Patton, R. A. Nutritional deficiency and sound-induced convulsive seizures in the rat. *Ment. Health Bull.*, 1946, **24**, 16–20.
110. Patton, R. A. The incidence of middle-ear infection in rats susceptible to sound induced seizures. *Amer. Psychologist*, 1947, **2**, 320. (Abstract)
111. Patton, R. A. Purulent otitis media in albino rats susceptible to sound-induced seizures. *J. Psychol.*, 1947, **24**, 313–317.
112. Patton, R. A. Sound-induced convulsions in the hamster associated with magnesium deficiency. *J. comp. physiol. Psychol.*, 1947, **40**, 283–289.
113. Patton, R. A. Vitamin B-complex concentrates and the incidence of sound-induced seizures in young albino rats maintained on purified diets. *J. comp. physiol. Psychol.*, 1947, **40**, 323–332.
114. Patton, R. A. The experimental approach to convulsive seizures. *Ment. Health, Bull.*, 1947, **25**, 15.
115. Patton, R. A. Maternal nutritional deficiency and the incidence of sound-induced convulsions in young albino rats. *Proc. Amer. Psychopath. Assn.*, 1947, **36**, 204–209.
116. Patton, R. A. Abnormal behavior in animals. In C. P. Stone (Ed.), *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951, 479–498.
117. Patton, R. A., Russell, R. W., & Pierce, J. F. The effects of auditory stimulation on the electro-convulsive threshold. *Amer. Psychologist*, 1949, **4**, 233. (Abstract)
118. Patton, R. A., & Zabarenko, L. M. Further observations on the incidence of middle-ear infection and sensitivity to sound-induced convulsive seizures in young albino rats. Unpublished manuscript, Univer. of Pittsburgh. (Cited in C. P. Stone [Ed.], *Comparative psychology* [3rd Ed.]. New York: Prentice-Hall, 1951, p. 492.)
119. Pilgrim, F. J., & Patton, R. A. Reversibility of sound-induced convulsions associated with a pyridoxin deficiency. *Amer. Psychologist*, 1948, **3**, 359. (Abstract)
120. Pilgrim, F. J., & Patton, R. A. Production and reversal of sensitivity to sound-induced convulsions associated with a pyridoxin deficiency. *J. comp. physiol. Psychol.*, 1949, **42**, 422–426.
121. Rabe, P. L. The cumulative frustration effect in the audiogenic seizure syndrome of DBA mice. *J. genet. Psychol.*, 1952, **81**, 3–17.
122. Reiss, B. F. The theoretical bases of convulsive therapy in relation to animal experimentation. *J. Pers.*, 1948, **17**, 9–15.
123. Russell, R. W. Comparative psychology. In C. P. Stone (Ed.), *Ann. Rev. Psychol.*, 1954, **5**, 234–237.
124. Scott, J. P. The use as test material of inbred strains of mice having high frequencies of audiogenic seizures. *Science*, 1950, **111**, 583.
125. Shafer, J. N., & Meyer, D. R. Effects of intense light stimulation on sound-induced seizures in rats. *J. comp. physiol. Psychol.*, 1953, **46**, 305–306.
126. Shaw, W. A., Utecht, A. J., & Fullanger, E. A. The effect of auditory stimulation upon the immediate retention of a previously learned maze behavior in the albino rat. *J. comp. physiol. Psychol.*, 1953, **46**, 212–215.
127. Shohl, Jane. Effects of exposure to sound on discrimination performance in the rat. *Psychol. Monogr.*, 1951, **65**, No. 12 (Whole No. 329).

128. Stainbrook, E. The experimental induction of acute animal behavior disorders as a method in psychosomatic research. *Psychosom. Med.*, 1947, **9**, 256–259.

129. Stainbrook, E. Experimentally-induced convulsive reactions of laboratory rats. II. A comparative study of post-convulsive maze behavior. *J. gen. Psychol.*, 1948, **39**, 191–216.

130. Vicari, Emilia M. Establishment of differences in susceptibility to sound-induced seizure in various endocrinic types of mice. *Anat. Rec.*, 1947, **97**, 407.

131. Vicari, Emilia M. A study of the genetic constitution of fatal and non-fatal sound-induced seizures in seven strains of mice. *Genetics*, 1948, **33**, 128–129.

132. Vicari, Emilia M. Age threshold of gene controlled sound-induced convulsions in mice modified by endocrine action and associated with physiologic threshold. *Genetics*, 1948, **33**, 632.

133. Vicari, Emilia M. Thiouracil prevents death which normally follows convulsions in sonogenic mice. *Anat. Rec.*, 1949, **103**, 594.

134. Vicari, Emilia M. Genetic aspects of the age threshold of susceptibility to sound-induced convulsions in various strains of mice. *Genetics*, 1950, **35**, **137.**

135. Vicari, Emilia M. Fatal convulsive seizures in the DBA mouse strain. *J. Pyschol.*, 1951, **32**, 79–97.

136. Vicari, Emilia M. Effect of 6*n*-prophylthiouracil on lethal seizures in mice. *Proc. Soc. exp. Biol., N.Y.*, 1951, **78**, 744–756.

137. Vicari, Emilia M. Hereditary fatal convulsions prevented by steroid hormones. *Genetics*, 1952, **37**, 634.

138. Vicari, Emilia M., Tracy, A., & Jongbloed, A. Effect of epinephrine, glucose, and certain steroids on fatal convulsive seizures in mice. *Proc. Soc. exp. Biol., N.Y.*, 1952, **80**, 47–50.

139. Weiner, H. M., & Morgan, C. T. Effects of cortical lesions upon audiogenic seizures. *J. comp. Psychol.*, 1945, **38**, 199–208.

140. Williams, R. J., Berry, L. J., & Beerstecher, E., Jr. Individual metabolic patterns, alcoholism, and genetrophic diseases. *Proc. nat. Acad. Sci., Wash.*, 1949, **35**, 265–271.

141. Witt, G., & Hall, C. S. The genetics of audiogenic seizure in the house mouse. *Amer. Psychologist*, 1947, **2**, 324. (Abstract)

142. Witt, G., & Hall, C. S. The genetics of audiogenic seizure in the house mouse. *J. comp. physiol. Psychol.*, 1949, **42**, 58–63.

143. Yeakel, Eleanor H., Shenkin, H. A., Rothballer, A. B., & McCann, S. M. Blood pressure of rats subjected to auditory stimulation. *Amer. J. Physiol.*, 1948, **155**, 118–127.

144. Zener, K. Personal communication.

145. Death rings the bell. *Science Illustrated*, 1947, **2**, No. 3, 33.

Received August 2, 1954.

Psychological Bulletin
Vol. 52, No. 6, 1955

SOME PERSPECTIVES ON "THE ATTENUATION PARADOX IN TEST THEORY"

FREDERIC M. LORD
Educational Testing Service

Dr. Loevinger (5) has performed a real service in focusing attention on certain of the earlier results (e.g., **1**, **3**, **4**, **12**) appearing in the literature regarding the relation of test "validity" to the item intercorrelations and to the distribution of item difficulty. The present note is intended to call attention to four known results that provide further clarification of the so-called paradox and of its effect on the selection of items in test construction.

Loevinger's "paradox" centers about the fact that, in a peaked test (a test composed of items all of equal difficulty), increasing the item intercorrelations beyond a certain point will lead to a decrease in a certain coefficient to which several writers have attached the term "validity." "Validity," in this special sense, is the product-moment correlation of test score with the common factor of the test items.

CURVILINEAR VALIDITY

As Brogden pointed out (**1**, p. 207), and as has been discussed elsewhere in some detail (**6**, **7**, **10**), the relation of test score to the common factor of the test items is necessarily curvilinear. This curvilinearity is commonly negligible for most tests in ordinary use, but it is not negligible (**6**, pp. 19–21) in the "region of paradox" with which we are here in part concerned—the region where the item intercorrelations are extraordinarily high. Consequently, in this region, the so-called "validity" coefficient is not an adequate measure of the virtue of the test. Obviously the use of the usual product-moment correlation coefficient tends to underestimate the true degree of relationship whenever the relation is curvilinear. What is required is a coefficient representing the curvilinear correlation between test score and the common factor of the test items.

The writer (**7**) reported a study, parallel to Brogden's, showing the effect of item intercorrelation and of distribution of item difficulty on the curvilinear correlation of the test score on the common factor. The results showed that no paradox existed when this coefficient was used to evaluate the test. However, as the writer pointed out, it would have been preferable to consider the curvilinear correlation of common factor on test score. This was not done because of the severe computational difficulties.

Brogden (**2**, p. 378) subsequently suggested in this connection that the "paradox" would have been found to exist if the preferable curvilinear coefficient had been used. Tucker[1] has recently shown it to be obvious that in the extreme case of perfect item intercorrelations, the curvilinear correlation of common factor on test score will necessarily be reduced by peaking the test. Thus the "paradox" does exist, even when an appropriate measure is used for the "validity" of the test over the whole group of examinees.

Cronbach and Warrington, in one section of their study (**3**, pp. 145–

---

[1] Personal communication, February 1954.

146) actually computed the preferable type of curvilinear correlation (η) for most of their tests. Their results confirm the existence of the "paradox." When η is used, the peaked test is shown to be preferable over the unpeaked test for higher values of the item intercorrelations than would have appeared to be the case if the inadequate product-moment correlation had been taken as the criterion.

DOES THE PARADOX ARISE IN COMMON PRACTICE?

A possible misinterpretation (in no way attributable to Loevinger) of Cronbach and Warrington's important article (*op. cit.*) will now be considered. These authors present results for five different levels of item "precision," denoted by σ_d. It is not always realized that only the last of these ($\sigma_d = 2.0$) lies comfortably within the range where aptitude- and achievement-test items commonly occur. Cronbach and Warrington give a table showing the interitem tetrachoric and phi coefficients corresponding to their values of σ_d; however, many test writers do not commonly deal with either of these coefficients.

On the assumption that all items are free-response items of 50-per cent difficulty and that all item intercorrelations are the same, Table 1 shows for each value of σ_d the corresponding value of (*a*) the reliability coefficient (r_{tt}) of a 100-item test (the split-half and the usual Kuder-Richardson coefficients are all equal when the items are statistically equivalent, as here assumed), (*b*) the tetrachoric item intercorrelation (r_{ij}'), (*c*) the interitem phi coefficient (r_{ij}), (*d*) the biserial correlation (R_i) between item and common factor, (*e*) the point-biserial correlation (r_{ic}) between item and common factor. Some of the less well-known relations between these coefficients for the present case are as follows:

$$\sigma_d = \sqrt{1 - R_i^2}/R_i,$$

$$r_{ij}' = R_i^2,$$

$$r_{ij} = \frac{2}{\pi} \sin^{-1} r_{ij}',$$

the last formula requiring the use of radian measure.

A glance at the reliability coefficients in the first row in the body of Table 1 shows that we seldom deal with tests having a value of σ_d much

TABLE 1

RELATION OF σ_d TO TEST RELIABILITY AND TO THE CORRELATIONS OF THE ITEMS WITH EACH OTHER AND WITH THEIR COMMON FACTOR

	Value of σ_d				
	0	.2	.5	1.0	2.0
Reliability of a 100-item test (r_{tt})	1.00	.998	.993	.980	.936
Item tetrachoric intercorrelation (r_{ij}')	1.00	.96	.80	.50	.20
Item fourfold point intercorrelation (phi) (r_{ij})	1.00	.82	.59	.33	.13
Point-biserial correlation between item and common factor (r_{ic})	.80	.78	.71	.56	.36
Biserial correlation between items and common factor (R_i')	1.00	.98	.89	.71	.45

smaller than 2.0. (Readers who are accustomed to working with item-test correlations may correctly assume that the last two item-common-factor correlations in the last two rows of the table are virtually the same in numerical value as the corresponding familiar item-test correlations.)

Cronbach and Warrington's data for η show that "paradoxes" occur, even in peaked tests, only when σ_d is less than about .75. It is thus seen that the paradox does not often appear in ordinary work with aptitude and achievement tests.

Is "VALIDITY" A BASIC CONCEPT?

Loevinger asks if it is not intuitively valid to demand that the basic concepts of psychometrics shall have a nonparadoxical relation to "validity." The present writer suggests that any over-all coefficient of validity for an entire group, even if properly measured so as to take curvilinear relationships into account, is from a theoretical point of view a mere hodgepodge, at least in the "region of paradox." One reason for this statement is the serious lack of homoscedasticity in the scatterplot that corresponds to the "validity" coefficient. When the item intercorrelations are high, the standard error of measurement is very different for test scores at different ability levels (**11**; **6**, pp. 13–16; **10**, pp. 533–536; **8**).

A more basic concept than validity is the discriminating power of the test at various ability levels. From the viewpoint of the mathematical statistician, a test score is a sample statistic that is to be used to test some hypothesis or (what amounts to the same thing) to reach some decision. The crudely stated hypothesis may be, for example, that the examinee is suitable for college work. The examiner has set up some more or less arbitrary rule for decision-making utilizing the test score as a guide. If the problem is properly formulated, all of modern theory of decision-making and theory of estimation and testing hypotheses can be brought to bear. It is possible, for example, to compare various psychological tests in respect to the statistical power of any specified decision procedure. An illustration of this approach is given in (**8**).

The discriminating power of any hypothetical psychological test with respect to any specified simple decision regarding any single examinee can be determined theoretically. An obvious index of this discriminating power would be simply the probability of reaching a correct decision regarding the given examinee (**8**, section 6). Another good index, in accordance with modern statistical theory, is the standard error of the maximum-likelihood estimate of the examinee's ability (**8**, sections 3 and 4).

It is too much to ask that the different discriminating powers of a single test for a wide variety of examinees and with respect to various unspecified decision problems should all be harmoniously and nonparadoxically summarized in a single average "validity" coefficient. Actually, if anything at all is known about the use to which a test will be put, it is usually apparent that different degrees of discriminating power are needed for examinees at different levels of ability. The test constructor's goal should be to achieve these various desired degrees of discriminating power rather than to maximize any single composite validity coefficient.

In actual practice, the discriminating power of a test at different ability levels might well be summarized by a diagram, as in Figure 1, where the

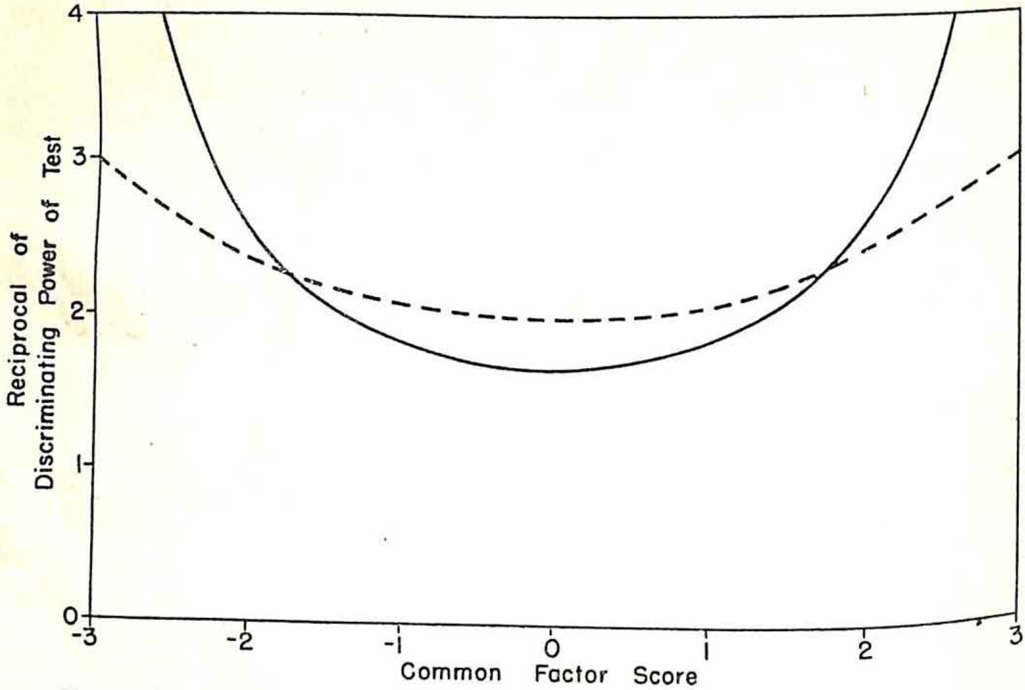


FIG. 1. THE DISCRIMINATING POWER OF TWO HYPOTHETICAL TESTS AS A FUNCTION OF THE TRAIT UNDERLYING THE TEST

ordinate is the reciprocal of the discriminating power. Here, the solid line represents a peaked test with items of 50 per cent difficulty and $R_i = .60$. The numerical values for plotting the curve were obtained from Table 1 of (8). The dotted line represents a (hypothetical) unpeaked test. The test constructor must decide in each instance what distribution of discriminating power best suits his purpose. Improved practical methods for preparing such curves for actual tests are in the process of development.

A MISCONCEPTION REGARDING PEAKING

Deplorably little has been written regarding the discriminating power of a test at specified ability levels. The present section is included in order to illustrate the theoretical importance of this concept, at the same time dispelling a rather serious misconception.

It is known that if a test is to be used only for discriminating between successful and unsuccessful examinees, then all items should be of the same difficulty level such that 50 per cent of the examinees near the cutting score answer each item correctly. This conclusion holds no matter how high the item intercorrelations are, and irrespective of any "paradox."

The misconception (not attributable to Loevinger) is that for optimum results the percentage of correct answers to each item should equal the percentage of examinees to be selected. Actually, (6, pp. 26–27) this misconceived rule can be correct only if the percentage of examinees to be selected is 50, or if each item is perfectly correlated with every other. Since interitem correlations (phi coefficients) are commonly nearer to .10 than 1.0, this rule will often be quite inadequate.

In order to determine the optimum

item difficulty for a test to be used for selecting a specified proportion, P, of the examinees, the following procedure is suggested (the procedure is outlined in **9**; the theory from which it obviously follows, in **6**, *loc. cit.*):

1. Convert P to a normal relative deviate, c.
2. Multiply c by the average item-test biserial correlation, denoting the result by h.
3. Find p, the area lying above the normal relative deviate h.

p is the desired optimum item difficulty (proportion of correct answers) for free-response items. A correction for guessing should be made if multiple-choice items are to be used.

Table 2 presents the optimum values of p for $r_{ic}'=.30$ and $r_{ic}'=.60$, the values of r_{ic}' being chosen as slightly beyond the usual limits of the range of average item-test correlation coefficients for aptitude and achievement tests. At such levels, the item-test and item-common-factor correlations differ very little from each other. A comparison of the second or of the third column with the first displays the error that would be made by following the erroneous rule mentioned at the beginning of the present section.

TABLE 2

Optimum Item Difficulty for a Peaked Test Having Specified Item Biserials

Proportion of Examinees to Be Selected (P)	$r_{ic}'=.30$	.60
.001	.18	.03
.01	.24	.08
.05	.31	.16
.10	.35	.22
.20	.40	.31
.30	.44	.38
.40	.47	.44
.50	.50	.50
.60	.53	.56
.70	.56	.62
.80	.60	.69
.90	.65	.78

SUMMARY

Four points have been brought forward:

1. The usual product-moment "validity" coefficient is inadequate. A curvilinear correlation coefficient must be used. The "region of paradox" is still found when the correct coefficient is used, although its size is reduced.

2. A greatly magnified notion of the extent to which the "paradox" occurs in actual achievement and aptitude testing is frequently encountered. Some numerical values are tabled so as to clarify this point.

3. It is pointed out that any overall "validity" coefficient for a group of examinees can be no more than an average of heterogeneous information. It is not surprising if this inhomogeneity produces so-called paradoxes. Such a composite coefficient is not basic to psychometrics. The discriminating power of the test for a specified decision problem regarding a specified examinee is the truly basic concept. A very simple index for describing this discriminating power is defined.

4. One instance where a consideration of the discriminating power of a test at a particular ability level leads to some practical conclusions (irrespective of the "region of paradox") is mentioned, and a common misconception is corrected. A table is given showing the optimum item-difficulty level for the peaked tests required when the examiner's purpose is merely to classify a group of examinees as successful or unsuccessful.

REFERENCES

1. Brogden, H. E. Variation in the test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. *Psychometrika*, 1946, 11, 197–214.
2. Brogden, H. E. Statistical theory and research design. In C. P. Stone & Q. McNemar (Eds.) *Ann. Rev. Psychol.* Stanford, Calif.: Annual Reviews, 1954. Pp. 377–400.
3. Cronbach, L. J., & Warrington, W. G. Efficiency of multiple-choice tests as a function of spread of item difficulties. *Psychometrika*, 1952, **17**, 127–147.
4. Gulliksen, H. O. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, **10**, 79–91.
5. Loevinger, Jane. The attenuation paradox in test theory. *Psychol. Bull.*, 1954, **51**, 493–504.
6. Lord, F. M. A theory of test scores. *Psychomet. Monogr.*, 1952, No. 7.
7. Lord, F. M. The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika*, 1952, **17**, 181–194.
8. Lord, F. M. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 1953, **18**, 57–76.
9. Lord, F. M. Optimum level of item difficulty. *Res. Memorandum* 53–3. Princeton, N. J.: Educational Testing Service, 1953. (Dittoed.)
10. Lord, F. M. The relation of test score to the trait underlying the test. *Educ. psychol. Measmt*, 1953, **13**, 517–549.
11. Mollenkopf, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189–229.
12. Tucker, L. R. Maximum validity of a test with equivalent items. *Psychometrika*, 1946, **11**, 1–13.

Received October 21, 1954.

Psychological Bulletin
Vol. 52, No. 6, 1955

SPECIAL REVIEW

SOME RECENT TEXTS IN EDUCATIONAL PSYCHOLOGY

DONALD SNYGG
State University of New York Teachers College

Blair, Glenn M., Jones, R. Stewart, & Simpson, Ray H. *Educational Psychology.* New York: Macmillan, 1954. Pp. xvii+601. $4.75.

Commins, W. D., & Fagin, Barry. *Principles of educational psychology.* (2nd Ed.) New York: Ronald Press, 1954. Pp. xvi+795. $5.75.

Cronbach, Lee J., *Educational psychology.* New York: Harcourt, Brace, 1954. Pp. xxvii+628. $5.50.

Mursell, James L., *Psychology for modern education.* New York: W. W. Norton, 1952. Pp. x+610. $3.90.

Remmers, H. H., Ryden, Einar R., & Morgan, Clellen L. *Introduction to educational psychology.* New York: Harpers, 1954. ix+435. $4.00.

Smith, Henry P., *Psychology in teaching.* New York: Prentice-Hall, 1954. Pp. xiii+466. $4.95.

Sorenson, Herbert, *Psychology in education.* (3rd Ed.) New York: McGraw-Hill, 1954. Pp. x+577. $5.00.

All psychologists have a personal stake in educational psychology. In the first place, the status and prestige of psychology as a profession depend to an uncomfortable degree on its effectiveness in the field of education. Almost all candidates for teaching certificates are required to take at least one course in educational psychology and a large percentage of them take no other psychology courses. To most members of this large and influential group, psychology means the educational psychology course they took in college and their attitudes toward psychology or toward psychologists will depend on the effectiveness or ineffectiveness of this course. The fact that educational psychology courses are often taught by nonpsychologists will have no bearing on the matter since people almost universally assume that anyone who teaches a psychology course is a psychologist. If he is not a *professional psychologist,* the whole burden of representing psychology as a profession falls on the text.

In the second place, educational psychology is important to psychology as a proving ground for theory and as a source of new concepts. The educational psychologist who fulfills his function by actually working on educational problems is required to deal with problems which can be ignored or postponed in the laboratory. The need for the educational psychologist to deal with a wider range of phenomena has always tended to make him more receptive to new concepts and has generally caused theorists in educational psychology to anticipate the concepts of the experimental laboratories. To the extent that educational psychologists extend their services to education, we should expect that their contribution to basic psychological theory will have to increase in the same way that clinical psychologists, in assuming more responsibility for clinical problems, have greatly increased their contributions to psychological theory. In fact, both of these areas are much more likely sources of fundamental theory than are experimental laboratories, which are best suited for testing it.

If this analysis is correct, anyone who wants to get a preview of future trends in psychological theory and, by way of the teaching profession, popular concepts of psychology and human nature, might do well to consider the content of current educational psychology tests.

Nature of the Individual

All seven texts focus their attention on the individual, who is regarded as a biological organism which is aware of itself and is actively, persistently, and more or less consciously seeking the satisfaction of certain basic needs, which are common to all people in all societies. Cronbach's statement (p. 100), that "Whenever a person acts, he is presumably trying to satisfy a need," is representative.

Behavior is molar and, in keeping with the general concept of activity from within, the term "stimulus" appears in the index of only one of the seven texts and is used there to refer the reader to a section critical of the S-R approach.

There is little disagreement about the nature of the basic needs, which are generally classified as physical (physiological) and psychological (in some versions, social or "socially derived"). Blair, Jones, and Simpson, along with Mursell, subdivide the psychological needs into social needs and ego needs. Cronbach points up the empirical nature of the specific needs cited by the different authors by denying the necessity of postulating a "need for security" by pointing out that "insecurity" is simply the fear of failing to satisfy another need. There seems to be a tendency to pull back and avoid the confusion that would result from compiling a long list of competing "basic" needs. The multiple needs concept badly needs a principle for predicting which need will operate when a number of needs are unsatisfied. Commins and Fagin handle this problem by compressing all of the needs into one and treating them as aspects of one basic need for self-maintenance and enhancement.

Needs may or may not be conscious. They build up tension and make the organism sensitive to those aspects of the environment which promise possible satisfactions. "Stimuli function only in the presence of needs" (Mursell, p. 48). Goals are derived only to satisfy needs.

Educational Implications

The dynamic nature of the individual postulated by the needs concept implies that psychology has a role in determining the goals of education as well as in assisting in the improvement of methods. If the needs concept is correct, educational goals and practices not consonant with human needs are unfruitful and probably disruptive. The object of education, from this point of view, is the satisfaction of individual needs or, in a large social context, the promotion of mental health.

Heredity and Environment

The emphasis on basic human nature and on the body as a biological organism have brought about a withdrawal from the extreme environmentalist position held by many educational psychologists a few years ago. The physical structure of the organism is accepted as an obvious determinant of mental functions, thus providing a role for heredity as a determinant of behavior.

Learning

A trend toward a cognitive point of view has enabled the writers who have pursued it most to bring perception, learning, problem solving, remembering, and personality development together as a single process and has greatly increased the number of pages devoted to this "learning" process in their books. Blair, Jones, and Simpson devote 229 pages to learning, Commins and Fagin, 292, Cronbach, 183, and Mursell, 211. In contrast Remmers, Ryden, and Morgan give learning only 52 pages, Smith, 143 pages, and Sorenson, 88 pages, treating many of the topics discussed by the first four sets of authors under other headings than learning. All of the first four give versions of the learning process which stress perceptual reorganization and can be summarized as follows:

1. Learner's awareness of need or goal.
2. Existence of an obstacle.
3. This requires learner to interpret the situation.
4. Attack on the problem in terms of the learner's interpretation.
5. Consequences of learner's attack.
6. Reinterpretation of situation.
7. New attack.
8. Learning stops when learner is satisfied.

In this point of view both effect and frequency lose their place as independent causal factors. Since the learner stops reinterpreting the situation as soon as he is satisfied, no "reinforcement" is needed to stabilize the response. And, as Mursell points out, practice or repetition gives opportunities for reinterpretation but does not guarantee that it will take place. Even Sorenson, who approaches the problem from an eclectic point of view and makes frequency the keystone of his learning system, seems to approach this when he remarks that repetition is not effective without interest.

The two other texts in the eclectically oriented group stress cognitive aspects of learning in their discussions without completely adopting the cognitive approach. Remmers, Ryden, and Morgan discuss trial and error, practice, and reasoning in terms of the learner's perceptions and accept insight. They explain conditioning by reinforcement. Smith, without citing Thorndike, uses exercise, effect, and belonging, and is the only author to make frequency or exercise an independent factor. He is also the only writer to modify the needs concept by accepting habit as a source of motivation. Commins and Fagin, and Mursell, who take the most consistently cognitive approaches, define habits as routine ways of dealing with situations which are formed, not by repetition, but by discovering what are for the time being satisfactory ways of response. These "habits" will be used as long as the corresponding situation is viewed in the same light and they result in satisfaction to the individual.

Transfer

One consequence of the rapid social and technical changes of the last few years has been the recognition that any education aimed at the teaching of specific facts and skills is apt to leave the student obsolete soon after he graduates. Awareness of the schools' need to educate their students for the "unforseeable" (Cronbach) has revived interest in the problem of teaching for transfer of learning. The consensus of opinion is that transfer is best achieved by promoting understanding and making learning meaningful. "A concept becomes useful by being understood" (Mursell, p. 310).

In harmony with the cognitive-field approach implicit in most of the texts, the theory of transfer by identical elements is minimized by all authors except Smith, who does not discuss transfer. Remmers, Ryden, and Morgan and Cronbach retain them as one cause of transfer. Teachers are advised to help the student to form generalizations and productive attitudes, particularly toward the self, since self-attitudes affect behavior in all situations. Mursell points out that the student must want to transfer and Remmers, Ryden, and Morgan suggest that the student be taught to look for familliar landmarks.

Mental Health

All the texts accept the promotion of mental health as an important goal of education. There seems to be an increasing tendency to describe health and adjustment in terms of the personal perceptions of the individual, although social and behavioral criteria are not neglected. Commins and Fagin stress the need for stable perceptual organization; Cronbach, self-acceptance; Mursell includes a consciousness of respect by others. Remmers, Ryden, and Morgan state that the purpose of education is to help develop an individual "who is able to solve his problems successfully in his own viewpoint as well as in the view of others." The other three texts similarly say that mental health is achieved by the satisfaction of needs.

CONTRIBUTIONS FROM OTHER FIELDS

Physiology

One of the most striking things about these texts to readers who remember the frequent references to the nervous system in early educational psychologies is the absence of such references in this group. The only index listing of the nervous system (in Mursell) leads the reader to an explanation of why consideration of what happens there does not help us understand learning. Two of the books discuss growth of the brain as an aspect of maturation but without going into neurological details.

Social Sciences

If we can judge by the indices, anthropology and sociology have contributed much more to these books than any other fields outside educational psychology itself. Havighurst is cited in as many books (five) as Guthrie, Hull, McGeoch or Thorndike; and Allison Davis and Margaret Mead (four each) in as many as Gordon Allport or Carl Rogers. Only two psychologists, Hilgard (seven) and Lewin (six) are cited in more volumes.

Experimental Psychology

The point of view used in these texts owes much more to clinical and social psychology than to experimental or comparative. Having acquired a body of data in their own field, educational psychologists, if these authors are representative, have almost completely abandoned the old practice of extrapolating from laboratory data. In the hundreds of pages devoted to learning, B. F. Skinner is cited only three times, all in one text, and Spence is ignored by all. Even Tolman, very surprisingly in view of the theoretical climate of the texts, is referred to in only two of them. Similar neglect of the work of the Hanover Institute, Cantril, Harlow, and Tsai, all of whom have furnished abundant ammunition for the cognitive approach, indicates that the channels of communication between educational psychology and experimental psychology are not well travelled.

It may be objected that the small representation of experimental psychologists in these texts is due to the fact that they are written for teachers and not for psychologists. (Commins and Fagin seems to be the only one of the texts which is directed toward both.) It may be wise to keep a discussion of such sources at a minimum in the introductory, one semester courses. But the fact that the authors of so many outstanding texts have felt able to ignore the current work in experimental psychology may be a sign of trouble. This apparent rift between the traditional home of psychological theory and the field which we ordinarily think of as the major field for applying psychological theory may be a sign of healthy diversification or of impending dismemberment. In any case, the dissociation should not be allowed to become complete. And the primary responsibility for preventing further separation seems to belong to the experimental psychologists. As the situation now stands, the most popular problems for research in "pure" psychology are learning problems derived from an objective S-R approach. There is no doubt that this conceptual scheme offers the best opportunity for neat, nicely quantified experiments which are easy to plan and likely to give quantifiable results. The authors of these seven textbooks are, by their analysis of behavior, challenging this approach. Either explicitly or by implication, they are saying that the conceptual schemes that now hold the center of the stage in the laboratories are not, in their present stage of development, useful to them as educational psychologists, and that as a result they have had to develop and use others. In comparison with the educational psychologist, who, like the physician, is bound by a professional obligation to do his best for individuals now, the experimental psychologist can work for more remote goals. At the present time he has two alternatives, either one of which is consistent with his obligations to society and to his profession. He can continue to refine the objective approach, in the hope that with further investigation and elaboration it will become useful in the applied fields. Or he can tackle the job of testing, refining, and clarifying some of the cognitive-field concepts that are being used in the applied fields. Since it would not be well to bet all our money on one horse, it is to be hoped that all experimentalists do not make the same choice.

Standards of Evaluation

In the reports on the individual texts which follow it is assumed that in addition to possessing the usual characteristics of any good textbook, a perfect textbook in educational psychology would be:

1. *Systematic.* The conceptual scheme should be adequate to make the subject and the student's own observations meaningful.
2. *Related to education.* The implications of theory for educational practice must be explicit.
3. *Related to psychology.* Educational psychology should be treated as an aspect of psychology and add to the student's knowledge of psychology as well as of education.
4. *Well documented.* Both psychology and education should be presented as fields for research. Theories should be presented as theories, not as revealed truth.
5. *Self-contained.* It is not possible to assume that the student using the book will have had any previous courses in psychology.

It is obvious that each of these characteristics conflicts with others. A book focused directly on educational problems is likely to lack systematic coherence. A book which concentrates on the psychological background is apt to lose touch with the interests and needs of school teachers. In addition, educational psychology is taught at almost all levels of college and graduate school, and for either one or two semesters. A two-semester book which meets our criteria would be so highly integrated that it could not be cut down for a satisfactory one semester course by omitting chapters.

BLAIR, JONES, AND SIMPSON

This book is closely tied to classroom problems. The theoretical position is consistent without being extreme or obtrusive. The authors have succeeded in striking a judicious balance without confusing immature readers with too many pros and cons. Taking the group of texts here reviewed, this is the "typical" book. Because of its clear presentation and reading ease it is a particularly good prospect for one semester courses. The chapters on evaluation of the school and on the teacher as a person are particularly good.

COMMINS AND FAGIN

The cognitive-field approach which is implicit in the majority of books here reviewed is spelled out in detail by Commins and Fagin. This book is written for psychologists as well as for teachers. I have the feeling that this forthright and sophisticated volume is primarily a psychology text in which the authors take over the field of education just to demonstrate the effectiveness of their point of view. Anyone who wants to relate the popular needs theory to the main stream of psychological theory would do well to read this book. Their discussion of learning from the cognitive-field point of view is the best to date. In all of our criteria except relation to educational problems this book is very high. Its relative weakness in that area is a result of the same orientation toward psychology which makes it an important contribution to educational theory. It seems best suited for students who are experienced teachers.

CRONBACH

Cronbach's text has caused more people to give up their plans to write an educational psychology text than any book that has appeared in years. It is unobtrusively but consistently systematic, well documented, and closely tied to educational problems which are presented at both the philosophical and classroom levels. The book itself matches the text. There are 97 illustrations.

The contrast between Commins and Fagin and Cronbach is an interesting one. Point by point, the theoretical approach of the two books is identical. But while Commins and Fagin use it in a whole-hearted attempt to feed it back into psychology and conquer new fields there, Cronbach is so uninterested in his conceptual framework, except as a convenient way of dealing with educational problems, that his text is completely inoffensive to advocates of other points of view. He is attempting to sell a point of view in education but not in psychology. In large departments where the educational psychology course is taught by several people with differing points of view about psychology, this is an important advantage for Cronbach.

This book rates near the top in all criteria and has already become a standard of comparison for other educational psychologies. It does have defects. The many case studies are apt to be confused with one another. It is just too big and too thorough for a one semester introductory course, although upper classmen who have had introductory courses in human development can manage it in that time.

MURSELL

Mursell stands with Commins and Fagin as an uncompromising, all-out exponent of the cognitive-field approach. The difference is that, unlike them, he is primarily interested in education. This is a brilliant and original book, closely reasoned and well documented with few references from outside the field of education. As a result it is not likely to give its readers much information about other points of view and is not intended to do so. Mursell's insights are acute and the book deserves thorough reading by psychologists. The presentation is so clear and well organized that it is a possible book for a one semester course at the undergraduate level.

REMMERS, RYDEN, AND MORGAN

Of the seven books, this book is the one farthest removed from the systematic position of Commins and Fagin and of Mursell. The text is characterized by a minimum of theorizing. It stresses environmental causes of behavior more than other books in the group. Although the needs concept is used, several statements seem to imply that the individual is stirred to action only by environmental stress. One result of this point of view is an excellent summary of social class as it related to education. Another strong point is an unusually long and good section on tests, interpretation of scores, and measures of correlation.

SMITH

This is a competent and somewhat conventional text at the undergraduate level. Smith gives the student a little of everything, stirring up needs theory and habit in the same pot, but he presents it charmingly and at a low level of reading difficulty. Departments offering one semester lower division courses might well consider his text.

SORENSON

Sorenson's well-known text, now in its third edition, has the advantages and disadvantages of the other texts which do not have a tightly defined, systematic point of view. Its strong points are a comprehensive selection of materials, a very readable style, good illustrations and summaries, and a legible format. It has been one of the most popular texts and will continue to be.

STONE, CALVIN P. & MCNEMAR, QUINN (Eds.) *Annual review of psychology* Vols. 5 & 6. Stanford, Calif.: Annual Reviews, Inc. 1954 & 1955. Pp. lx+448 & vi+517. $7.00 (each Vol).

A reviewer setting out to review the *Annual Review* feels himself caught in the toils of an infinite regress. One way of escape lies in an over-all appraisal, rather than the more usual chapter by chapter treatment, but Lanier has already done this admirably for the first four volumes.[1] Perhaps the fact that the sample has now increased by 50 per cent may serve as an excuse for reopening some of the more general questions.

Lanier quotes from the preface to Volume 1 where it is stated that the authors were asked to "adopt an interpretative and evaluative approach to the literature they selected for review." He felt that in this respect Volumes 3 and 4 improved upon their predecessors, but still contained a number of chapters which were merely "bibliographic digests." An attempt was made in the present review to assess Volumes 5 and 6 using the same standards. Volume 5 seemed to regress slightly but Volume 6 was in this respect the best so far. Even here, however, evaluation and interpretation were not always conspicuous.

Instead of blaming the authors it may be well to ask whether an "interpretative and evaluative approach" is possible on an annual basis. Such an approach would seem to involve the drawing out of implications from experimental data, the restructuring of problems in the light of evidence from different areas, and above all the identification of trends. These things are not easy when time presses, and the last of them can hardly be done at all. Psychology, like the watched kettle of the proverb, changes so imperceptibly under this constant scrutiny that from one year to another it may not seem to change at all. The identification of trends is more appropriate to a quinquennial than to an annual publication.

Then again many readers may want a competent bibliographic digest. The research worker can go to *Psychological Abstracts*, but the busy teacher or clinician must often find in the *Annual Review* an invaluable clue to the labyrinth of the journal literature. So long as the material is complete and objectively presented, he is quite prepared to do his own evaluation.

Looking at the two volumes under discussion with this thought in mind some qualification appears to be necessary. Where the topic itself provides the organization, as for instance in the chapters on the special senses, the factual approach works admirably, and the relevance of the data presented is established by a fairly well-articulated context. Elsewhere, however, this may not be so, and a merely bibliographic treatment sometimes leaves lost little facts adrift on pages. For example, "In a sharing test the only child was more selfish, on the average, than the child with sibs, according to Ugurel-Semin."

Particular chapters, again, seem to justify the original terms of reference quoted above. For example, E. Lowell Kelly on "Theory and Techniques of Assessment" in 1954 and MacCorquodale on "Learning" in 1955 have written penetrating and objective surveys in which the interpretation and evaluation is of the highest value. On the other hand it must be kept in mind that although these chapters reported the literature in their fields and periods they were really much wider in their scope. Each contained a far-reaching appraisal of the present situation as well as sound counsel for the future. It will be some years before the same thing can be done again, and probably much longer before it can be done as well.

---

[1] Lanier, L. H. An evaluation of the *Annual Review of Psychology* (Vol. I–IV) *Psychol. Bull.*, 1954, 51, 180–190.

Other chapters justify their approach in other ways. Tolman and Postman's provocative discussion of papers on learning in the 1954 volume gains from its avowed bias, though here and there the writing is a little slapdash. "It would seem to the reviewers that nature must be even cleverer about providing the right environmental stimuli for combining movements into adjusted, directed acts than we would have supposed her capable of." Here and elsewhere the authors might have done a little more running back and forth in front of what they had written, but a coldly objective treatment, however grammatical, would lose by comparison with the present chapter. Another idiosyncratic chapter in the same volume is by George Miller on communication. He selected his papers "by rigid adherence to a single rule: Only those studies were covered that interested the reviewer." On this basis he has put together a lively, forward-looking discussion which might have been impossible had his bibliographic burdens rested more heavily upon his shoulders. On the other hand the danger of allowing the reviewer too free a hand, particularly in a central area, shows up in the 1955 chapter by Kallmann and Baroff on "Abnormalities of Behavior." Here the original title has had to be supplemented by "In the light of psychogenetic studies" presumably in order that it may represent more nearly what is in fact discussed.

The upshot of all this seems to be that it is possible to justify a sound objective survey of a year's research, particularly in the more structured areas, and it is possible to justify an "interpretative and evaluative approach." But some question remains as to the best way in which the two can be combined. Assuming that interpretation and evaluation can hardly be made afresh in each year, although they remain an essential element in the series as a whole, it might be worth while to envisage some more formal provision for them at longer intervals. A sort of rotation could no doubt be set up among the various fields which would permit each to lie fallow so far as the more systematic kind of treatment is concerned. In any given volume the authors charged with interpretation and evaluation might be allowed the time and space their task requires. The idiosyncratic chapters might be left to happen as happen they will.

Volume 6. In the concluding paragraph of their preface to this volume, the editors say: "When scientific psychology flourishes on a world-wide basis, as in a few nations of the western hemisphere it has developed during the past half century, the *Annual Review of Psychology* will fully realize its major aims: to disseminate the fruits of research in psychology and to promote collaboration among scholars of all nations. . . ." Leaving world wide plans to the future, it seems reasonable to ask how well the *Annual Review* is serving the "few nations of the western hemisphere" now. A study of the literature cited reveals some interesting differences between one area and another. Classifying the sources into "American" and "Foreign" we get the following figures:

	American	Foreign
Child Psychology	101	2
Learning	148	0
Vision	159	35
Hearing	117	23
Somesthesis and Chemical Senses	29 (36)	37 (18)
Individual Differences	75	3
Personality	58 (49)	26 (0)
Social Psychology	137	2
Industrial Psychology	116	6
Comparative Psychology	59	32
Physiological Psychology	119	31
Abnormalities (Psychogenetics)	96	48
Assessment	155	8
Psychotherapy	89	4
Counseling	78	0
Statistical Theory	84	20
Educational Psychology	109	3
Problem Solving and Thinking	161	4

Since the authors of the chapters on "Somesthesis and the Chemical Senses" and "Personality" were foreign, the corresponding figures from 1954 have been added to make inter-area comparison easier. It would seem that the sensory, physiological, genetic and statistical chapters resemble one another in having a considerably greater proportion of foreign references than do those with a more narrowly psychological content. Classifying along these lines and using the 1954 figures for the two chapters mentioned above, we get a mean foreign percentage of 34.2 for the first group of chapters and 2.73 for the second. The difference is significant, but what does it signify? There seems no reason to suppose that one group has more linguistic versatility than the other; in fact the chapters by Kallmann and Baroff and by Teuber have fewer foreign references than the average for the chapters with which they have been classified. The suggestion that there is less European work going on in the areas with low foreign reference scores seems plausible for learning theory and counseling which are to some extent American preoccupations, but this answer does not appear to be altogether adequate elsewhere. Nuttin found 26 European studies on Personality worth reporting in 1955 whereas Child in 1954 had found none. Similarly Heron's 1954 chapter on Industrial Psychology lists 33 European titles as compared with 6, all from one journal, included by Wallace and Weitz. Undoubtedly more work is going on in America than elsewhere, and a proportion of three to one seems reasonable. A proportion of thirty to one suggests that in many areas research reported overseas is being neglected. The significant difference referred to above is probably due to the fact that some areas are more highly integrated than others, and what is discovered anywhere is relevant everywhere. As in all established sciences a parochial attitude is incompatible with progress. In other areas data tend to accumulate piecemeal, and accessibility alone determines what shall be added to the heap.

On the other hand it is clear that, if the *Annual Review* is going to realize the aims laid down in the Preface to Vol. 6, some further thought must be given to this problem. Though to the reviewer difficulties of time and language seem insuperable, the impressive performance of some authors suggests that the level of aspiration might be raised for all.

Looking back over this review it seems that too much time has been spent trying to tell a body of able and energetic psychologists how to do better what they are already doing extremely well. The *Annual Review* fulfills an indispensable function in the psychological world, and the 1955 volume worthily continues a short but honorable tradition.

JAMES DREVER
University of Edinburgh

ALLPORT, FLOYD H. *Theories of perception and the concept of structure.* New York: Wiley, 1955. Pp. xxii+709. $8.00.

This book is in reality two books. The author planned it as one: the earlier chapters read as if preparatory for the later, as the later chapters purport to be consequences of the earlier. But the reader is altogether likely to find it the product of two basically different intellectual enterprises, which are well announced in the double title. The first twenty chapters give us a critical survey of many theories of perception while the last chapter presents the author's very original concept of "structure" which may be applied to, but does not necessarily arise out of, discussions of perception.

I. As a review of the many theories suggested by experimental research on perception, Allport's work is noteworthy. He has certainly done a conscientious job on each of the thirteen major theories—ranging from the Titchenerian core-and-context, to the cybernetics of Wiener and of Pitts and McCulloch, and including such variety as the Gestalt, the Hebb cell-assembly, Helson adaptation-level, Brunswik probabilistic, and the Bruner *et al.* directive-state. As each in its turn is taken up, the author is at great pains indeed to give it an adequate and detailed exposition —even occasionally at the expense of considerable restatement and/or meticulous and finespun argumentation. Each is a model of clarity; and many presentations of a technical system of thought with scarcely any diagram at all is an object-lesson in effective and precise and yet graceful use of the English language. It is, in fact, something of an intellectual treat to follow Allport, especially as his thoughts weave in and out and sweep backward and forward in utter command of the so diverse and often recondite theoretical systems. Incidentally, he does not burden the pages with pedantic chapter and page references; and less than three hundred works are cited in the course of the whole survey.

After exposition, criticism! Each field of experimental research and the theoretical system arising from it is examined and appraised; and where each falls short of meeting the complete demands of a perceptual theory is carefully (sometimes repetitively) pointed out. The reader frequently feels that he is an auditor at an Allport seminar on the Psychology of Perception. And he may recognize some of Allport's ways of thinking: his emphasis upon the close inductive relation between experimental finding and theoretical superstructure; his lifelong hardheadedness toward the enthusiasts for molar (and institutional) explanation who forget that, after all, without the parts you have no whole; his meticulous thoroughness in canvassing the implications of a viewpoint; his insistence upon the parsimonious explanation; his respect for logical consistency.

Toward the end of this survey the reader comes upon a baffling Chapter Nineteen in which is pointed up one defect in all the theories: "the unsolved problem of meaning." The present reader at least was led to anticipate a positive contribution to supply this lack. On the contrary he found himself in agreement with one of the authors cited (Pickford) in whose article he found this sentence: "The idea of meaning is a psychological abstraction, and it is necessary to guard against the risk of treating it as if it were a discrete entity."

Nevertheless, the present reviewer is of the opinion that for anyone about to work in the field of perception, this book will be a first recommendation, both for inclusiveness of the perspective and for lucidity of each portrayal.

II. Following this exhaustive survey of the thirteen major theories of perception Allport does not attempt to reduce them to one; instead he draws up some eight basic (and quite abstract) generalizations. The perceptual process has, in all the theories, had the character of internal relatedness, circularity or self-closedness, space and time building, flexibility or deformability, persistence of constant relationships, level-maintenance or equilibrizing, weighting and pooling of energies or dimensions, and interaggregate facilitation or opposition. These broad (and formal) conceptions give us our most substantial insight into the nature of perceptual process, over and beyond the basic fact that a perceptual act is really a dynamically operating structure.

Prior to and independently of his examination of the theories of perception Allport had been seeking generalization of all natural laws in the concept of "structure," as set over against number or dimension or quantitative description, not a spatial or a temporal structuring simply but an event-structuring, a pattern of "ongoings" and of "events." Ongoings include such things as neural impulses or muscular contractions (to employ only neural illustrations here). Events are "junction points" as when an ongoing stimulus-process hits a receptor or an ongoing neural process at a synapse fires a succeeding neuron. The enterprise the author is embarking on is what may properly be called a metaphysics—in the original and precise use of that term as connoting the effort to see the world in its most abstract and generalized aspects, with no implication at all of a sneaking behind the scenes but simply that of highest-level description.

To give a touch of concreteness to his highly abstract theory of event-structures, Allport suggests a simple mechanical model or analogue, a visualized schema. "Let us think of a thin wire hoop, capable of being bent in any direction or shape, but always remaining a hoop, upon which are placed, at intervals, a number of cross-marks . . . cutting the hoop into segments. The model is, of course, kinetic, not static. The segments of wire into which the hoop is broken by the cross-marks represent ongoing processes; the cross-marks are the events between these ongoings . . . 'event-points.' A succession of ongoings and events thus occurs around the hoop, and the succession can repeat itself indefinitely" (p. 634). With the one hoop we can conceptualize a complete and repetitive cycle of ongoings and events. With two such hoops at right angles to each other we can visualize an event-system. With a number of hoops put adjacent in a larger circle of hoops, each touching two others at event-points, we have a model for a hierarchy, or cycle of cycles of ongoings and events. The present reviewer will not follow the further elaborations of this exposition of key-concepts illustrated with this hoop model; the reader can do so by consulting the book. Suffice it to say that manipulations of the hoops furnish the author with visualized parallels for observed facts of organismic behavior, as well as for particular principles of configuration of learning, of equipotentiality and mass action, of equilibrium and steady-state, of set, of motivation, of constancy, of facilitation, and so on.

Unquestionably there is ingenuity manifested throughout all of this. Yet the present reader has experienced the repeated qualm that it is all highly artificial, that the architect of the hoops has devised them to schematize certain principles already observed, without obtaining any new insights into the phenomena themselves. If there is benefit to be derived from such purely analogical and artificial parallelisms we may expect to find it in any new hints and suggestions given to research workers. (After all, the farfetched diagramming of Lewin did lead to new envisagements of concrete research problems. And other cases are matters of history.)

Returning now from diagrams to principles, the question remains: how much will Allport's concept of structure clarify the problems of perception? This reviewer would leave it at this: in the same measure that his structural theory serves as a general integrating law for nature, and more particularly to the extent that the notion of "ongoings" and "events" turn out to prove fruitful to those who work on perceptual problems.

JOHN F. DASHIELL
The University of North Carolina

RÉVÉSZ, G., (Ed.) *Thinking and speaking: A symposium.* Amsterdam: North-Holland Publishing Co., 1954. Pp. 206. $4.00.

This book is a reprint of a double number of *Acta Psychologica* and contains ten contributions. Five are by psychologists (Eliasberg of New York, Kainz of Vienna, Piaget, Cohen of Manchester, and Révész), one is by the philosopher Jørgensen of Copenhagen, one by the mathematician Waerden of Zurich, two by psychiatrists (Goldstein of New York and Gruhle of Bonn), and one by the linguist Buyssens of Brussels. English summaries are provided for the six articles not in English. Révész states the problems: Is speech possible without thought? Is thought possible without speech? He answers the first question affirmatively. The second is more difficult, but he reaches the conclusion, after examining many kinds of thought, that thoughts that are *wortlos* are not *sprachlos*, but are without exception based on the language-function. The reason for this conclusion is that wordless thoughts during their development have a syntactical structure, an anticipated schema in Selz's meaning of these terms. Révész also emphasizes the importance for thinking of memories of events previously formulated in speech. He believes that he has proved that the two functions are one and indivisible (though speech may occur without thought). Piaget in a genetic study states that the two functions are "supported by each other and are in a perpetual interaction, but either depends on intelligence, which is anterior to and independent of speech." Cohen's paper is a thesaurus of opinions of past writers on these problems and also presents a preliminary experiment on the effect of delayed recording on the quality of thought, in which students were required to write imaginative stories on a set theme, starting to write either immediately or after a period of time in which they thought about the task. The other papers offer a great variety of detail about animal speech, aphasia, abstraction, the thinking of mathematicians, and language in relation to Goldstein's concept of self-realization. In an epilogue Révész says that these distinguished authorities have not come to an agreement, but their efforts have led to important conclusions and have cleared up many misunderstandings and paradoxes.

This book should be read by all American psychologists who are interested in keeping abreast of the developments of *Denkpsychologie*. It seems somewhat foreign to the atmosphere of American psychology, with its preoccupation with newer problems than this one which seem to offer more hope of solution by other means than discussions of this kind, however excellent they may be.

FRANK A. PATTIE
University of Kentucky

Handbook of human engineering data. Port Washington, N. Y.: Special Devices Center, Office of Naval Research, 1952. (2nd Ed. Rev.) Pp. 607. $9.00.

This *Handbook* was developed by the Institute for Applied Experimental Psychology of Tufts College, Medford, Massachusetts, under contract with the Special Devices Center of the Office of Naval Research. Work on it was begun in 1947. After three years of work by the Tufts Handbook Staff and a large number of collaborators over the country, the first edition was issued in loose-leaf form in December, 1949, not as a finished product, but rather as a step in achieving a more complete reference work. This was the first major publication of its kind to appear in the newly constituted field of human engineering, and like any pioneering effort, it faced many difficult problems of the type which usually require a stepwise solution.

The Second Edition of this *Handbook*, issued in August, 1951, was a combination of the original plus Supplement No. 1. Supplement No. 1 included an entirely new Part IX, *Learning*, and a section on *Altitude* fitting into the already existent Part VII, *Physiological Conditions as Determinants of Efficiency*.

The Second Edition (Revised) differs from the Second Edition in the replacement of the original Part VI, *Motor Responses* with a completely rewritten Part VI, *Motor Performance*, including the associated indexes and glossary, which is known as Supplement No. 2. The essential features and plan of the original edition have been adhered to throughout all of the releases. This includes a 22-page glossary and an author and subject index at the end.

Two major problems which a reference work of this type faces are the intended audience, and the rationale of organization and presentation of materials. Who is going to use the book, how, for what, how frequently and under what circumstances? Now these are obviously not new problems; they are simply of special significance because of the nature of human engineering information and design engineers, who were originally conceived as *the* audience for the *Handbook*. A fine job of literature reviewing and of condensation of information about human beings has been done by the Tufts College staff. It turns out that the manner and method of presentation and organization is not so felicitous from the point of view of the intended engineer users. This is not entirely true because there is a small percentage of engineers who have been aware long enough of the fact that they must consider "user" characteristics if their design work is to be successful, and have done something about it, to have developed some understanding of pertinent data. The great majority, however, appear to need "answers" initially without personal study of an additional specialty (to them). The sophistication required for effective reading of the *Handbook* and for ready access to desired subject matter therefore renders it less useful to the large majority of engineers than is to be desired. However, by way of compensation, the *Handbook* has been found to be an effective reference by workers in several other fields such as engineering psychology and other phases of human engineering, military personnel, training activities, and business establishments, to mention only a few.

To many nonpsychologists the *Handbook* provides the means whereby they first become aware of the broad panorama of scientific concern with man in a nonclinical sense. This reviewer also believes there should be a growing place for the *Handbook* in training programs throughout the country, whether in college, government, or industry.

Much might be said about the problems of display of information and layout of the book, but particular solutions to these problems are so intimately associated with content and intent that perhaps the only things worth saying in a review are that (*a*) there is more emphasis on giving information on human beings than on recommending engineering practices which indicate effective use of the data, (*b*) the organization is based upon human topics, not engineering ones, so that engineers report having difficulty finding the data relevant to their particular problems. Be that as it may, it is hardly to be expected that a single reference handbook could be prepared that would meet all of the needs of design engineers and of the motley field of human engineering at one fell swoop! The book already meets important needs and offers promise of meeting others as already indicated above. This will take place as more nonpsychologists become aware of the need for considering human factors *systematically* in making things for people to use or operate, and in modifying the environment in which men must work or live.

The *Handbook of Human Engineering Data* in the opinion of this reviewer is an important contribution which, teamed up with two other major publications in human engineering—Chapanis, Garner, and Morgan's *Applied Experimental Psychology* (Wiley, 1949) and Woodson's *Human Engineering Guide for Equipment Designers* (University of Calif. Press, 1954)—provides a strong basis for growth and effective practice in human engineering. Consequently people interested therein should have the *Handbook* available to them.

Arnold M. Small
U. S. Navy Electronics Laboratory
San Diego, California

McNemar, Quinn. *Psychological statistics.* (2nd Ed.) New York: Wiley, 1955. Pp. vii+408. $6.00.

This edition maintains the useful pedagogical device of stressing algebraic derivation as an aid to understanding of formulas. Teachers and students who liked the content and style of the first edition should like the new edition even better since it is an improved arrangement and elaboration of the 1949 version.

Because the original edition is so well known it suffices to note some of the topics which have been clarified or added. Much fuller discussion is provided of types of error and power considerations in hypothesis testing as well as the problem of designating one-tailed vs. two-tailed critical regions. Among the "significance tests" which were not presented in the first edition are chi square for several correlated proportions, differences for related variances, differences for related correlations, Bartlett's test for homogeneity of independent variances, exact probability for the 2×2 table, and a reference to the comparison of means when variances are unequal. Brief attention is also devoted to intraclass correlation, the discriminant function and Kendall's tau while a more friendly attitude is taken toward the point biserial and the phi coefficient.

The most extensive revision has been made in the treatment of analysis of variance, which presents an excellent discussion of underlying mathematical models and their relation to choice of error term for F tests. In a new section on latin square, McNemar offers a strong argument for limiting use of this design in psychological research, but does not mention the experimental advantages of replicated latin squares. In view of recent trends and his conservative attitude toward Type 1 errors, it is of interest that McNemar still advocates use of the t test for comparing individual means following a significant over-all F. Since a section is devoted to the handling of unequal frequencies in the single classification design, it would have been appropriate if reference had also been made to the common problem of disproportionate subclass frequencies in multiple classification design.

A new chapter on nonparametric methods contains four pages. As McNemar notes in the preface, "Some will be critical of Chapter 18 because it does not contain all the several so-called nonparametric techniques."

It is superfluous to add that this text is a very solid introduction to descriptive and inferential statistics. The reviewer understands that the first edition was widely adopted. Research will continue to benefit as the new edition is mastered by students of psychology.

LEONARD S. KOGAN
Institute of Welfare Research
Community Service Society of
New York

PIAGET, JEAN. *The construction of reality in the child.* Translated by Margaret Cook, New York: Basic Books, 1954. Pp. xiii+386. $6.00.

The way in which the infant organizes the universe of objects in space is analyzed from observations and experiments made by Piaget on his son and two daughters from birth to 18 months. To elicit responses, objects were moved, dropped, hidden behind screens, partially exposed, moved so that part of the trajectory was hidden, etc.

In the first two stages (up to 6 months), the infant's universe is not cut up into objects but is like a stream of pictures without substantial permanence or spatial organization. But every intersensory coordination, every movement, and every functional use of sucking, sight, hearing, and touch contribute to the anticipations which assure the solidity and coherence of the external world. If the new baby had object-concepts, he would search for vanished objects and remove obstacles. But he does not conceive the vanished image as an object which exists in space and remains identical with itself.

In the third stage, if the object is not found within the exact extension of his movement, the child gives up hope of finding it. When prehension comes, the child learns to follow with his hand objects which escape him and thus begins to attribute permanence to tactile objects.

In the fourth stage, the child searches with his hands for the dis-appearing object without reference to the displacement. When he glimpses any part of the object he shows greater desire to see it than when he catches sight of the whole. He is trying less to free an object masked by a screen than to free his own perception.

In the fifth stage (13 to 18 months) permanence is attributed, since the child not only expects to find vanished objects in the places where they were left but also in the extensions of their trajectories.

In the sixth stage, the child can construct invisible objects by means of mobile schemata which can be combined in various ways. Thus he discovers the object's real permanence and constructs a collective universe.

The development of the spatial field begins in the first two stages with heterogeneous and practical groups in which each perceptual bundle constitutes a space. In the third stage practical groups are coordinated into subjective groups, while in the fourth stage reversible operations are discovered and a transition made from subjective to "objective" groups. This becomes complete in the fifth stage, which makes possible a sixth stage in which the child can, by the representation of movements, rediscover a hidden object after several sequential displacements, even if some occur outside the visual field.

Similarly during the first two stages of the development of causality, the child makes contact between internal activity and the external environment by means of pure reflexes and elementary habits. The third stage, that of magico-phenomenalistic causality, begins with the coordination of prehension and sight into secondary circular reactions. Elementary permanence is attributed to things as a function of the action, and systematic interest in causal relations appears. For example, the child who sees his hands and feet move in the visual field and who already controls movements of his hands to some degree, comes to their causal connection which is so clear to the observer but not to the child. In the fourth stage the child begins to externalize and objectify causality; in the fifth stage real objectivation and spatialization become evident. In the sixth the universe has come to be an independent system of causes and effects, among which the sequences of his own acts are placed as elements in a totality which transcends them.

In general the organization of the temporal field parallels that of space and complements that of objects and causality. In the first two stages there are practical time relations, since the child knows how to coordinate his movements in time and to perform certain acts in a regular order. But he is unconscious of the unfolding. In the third stage the child begins to act upon things and to use their interrelations through prehension of objects. Because these movements depend upon the personal action of the child they are called subjective. Secondary circular reactions appear which in the fourth stage result in schemata which make possible application of familiar means to new situations, and thus lead to the objective series of the fifth stage in which time is applied to things and a continuous and systematic linkage formed which unites the events of the external world to one another. In the sixth stage the child goes beyond the present by evoking images or language signs which constitute the representative series.

Three formative processes are necessary to the elaboration of the object concept; the accommodation of organs which makes possible the foreseeing of the appearance of bodies; the coordination of schemata which makes possible the endowment of each of these bodies with many interconnected qualities; and a form of sensorimotor deduction which makes possible the understanding of the displacements of bodies and the reconciliation of their permanence with their apparent variations. These functional factors—foresight, coordination, and deduction—change entirely in structure when they pass from the sensorimotor plane to that of speech and conceptual operations, when systems of classes and thoughtful relations are substituted for simple practical schemata.

Thus in his first eighteen months the child proceeds from an initial practical solipsism to the construction of a universe in which he is an element in a stable world. In the initial state the universe is neither substantial nor extended in depth as it is related to a subject ignorant of himself and perceiving reality only through his own activity. The final stage is that of a solid and vast world obeying the laws of conservation (objects) and kinematic ones (groups) in which the subject places himself consciously as an element. From egocentrism to objective relativism seems to be the formula of this law of evolution.

JOHN E. ANDERSON
University of Minnesota

ROGERS, CARL R. & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. viii+447. $6.00.

This is a truly impressive book, if one holds any brief for the value of objective research where psychotherapy is concerned. It is still much too early to hope that discussions of psychotherapy can be aimed at concrete problems involving specific principles. Practical reality still demands the teaching of untried principles spiced with a bit of emotion to help convince the apprentice that he is getting his money's worth. This progress report from the Counseling Center of the University of Chicago should, however, convince anyone that objective research on psychotherapy is possible and may, eventually, pay off with concrete results.

It should also convince the fainthearted that research on psychotherapy with real patients who have real problems might best be left to those who have great courage, considerable dedication, and the foresight to equip themselves in advance with a large grant from a foundation. This kind of programmatic research in psychotherapy cannot substitute for the brilliant hunch, but the brilliant hunch must inevitably be subjected to the torture of programmatic research.

Twelve members of the staff of the Counseling Center report separate but interrelated studies of the same group of twenty-nine unselected clients who received client-centered psychotherapy. As the studies were designed to evaluate the outcomes of this particular brand of treatment, two simple but, from the practical standpoint, quite difficult control procedures were devised. Half of the clients receiving therapy were asked to wait sixty days before starting treatment in order to determine if expressed desire for treatment produced changes in adjustment. The results were negative. The second control consisted of a matched group of persons who had volunteered to participate in research on personality. Although the second control group did not receive treatment, they were subjected to the same tests as the therapy group in order to determine whether time produces changes in personality. The results again were negative.

The simplicity of such a design belies, of course, the efforts necessary to bring to fruition a four-year study of eighty subjects and sixteen therapists. The measuring instruments, moreover, present the same problems of validity and reliability in a carefully worked-out research design as they do in the diagnostic clinic. In line with Rogers' interest in the self concept as an explanatory construct in personality change, major interest centered on Q sorts. In addition, the Thematic Apperception Test, the Willoughby Emotional Maturity Scale, attitude scales, therapist ratings, friends' ratings, a situational test, and phonographic recordings of all treatment interviews, comprised the data for the objective studies. The six-hour battery of tests was administered to clients and controls twice before treatment of the clients was begun, again at the conclusion of therapy, and, finally, six months to a year following the end of treatment. The Q sorts (used to compare self, self ideal, and the ordinary person) were administered several times during therapy depending upon the length of treatment. Two follow-up interviews and a follow-up questionnaire completed the pattern. All analyses were based upon carefully worked-out hypotheses which were congruent with some aspect of personality theory as viewed by the Chicago group.

The results are fairly clear-cut in showing that the therapy group underwent significant changes in their attitudes toward themselves as measured by Q sorts. These changes were not found in the control groups. Furthermore, the post-theory concepts of self in the therapy group approached more closely the clients' self-ideals and their conceptions of "other" people. The latter two concepts remained relatively constant. During the follow-up period, regressive changes took place in some of the therapy group whereas other members held their gains.

The results of the Q sorts were, in general, upheld by analyses of the other measures utilized. Of particular interest, because of its novelty, is the analysis of friends' ratings of behavior. The friends were kept in ignorance of the therapy. For clients who were judged to be making progress in therapy, the friends' ratings indicated a definite increase in maturity of behavior which was not found in friends' ratings of clients who were not improving. For once, therapy research combined measures of self-evaluation with external ratings of behavior in *normal* social situations.

With the more intensive analysis of the phenomenal field which this book represents, the expected perplexing problems arise. A positive relationship seems to exist between *negative* self-evaluation and internal tension. However, where *positive* self-evaluations are concerned, the interpretation is not so clear, as the positive statements may be valid or they may represent defensiveness. Such a possibility, of course, raises considerable question about the validity of any method which depends on positive self-evaluations as indicators of change in personality integration. The dilemma may not be insoluble.

This book is another testimonial to the fact that psychotherapy is rapidly becoming a legitimate field of scientific research as well as an applied art.

VICTOR RAIMY
University of Colorado

GRINGS, WILLIAM W. *Laboratory instrumentation in psychology*. Palo Alto, Calif.: The National Press, 1954. Pp. vi+282. $4.75 (cloth), $4.00 (paper).

In many areas of psychological investigation, it is obvious that progress is dependent upon developments in technological fields far removed from the introspectionist's armchair. It is equally clear that the meaning of data gathered with the help of any gadget more complex than a pencil is relative to the instrumentation. Appreciation of the possibilities of use and misuse is a prerequisite for both competent investigation and critical evaluation. And so the techniques course has become an increasingly familiar component of graduate training, and psychologists of older vintage often find themselves in need of informal tutoring.

It has been Grings's purpose to provide in convenient form a discussion of the basic characteristics of representative stimulating and recording systems, principally for use with human subjects. The material is organized in seven major chapters—Behavior Recording Systems, Timing and Counting, Audition, Vision, Other Senses, Human Learning and Perception, and Bioelectricity—and is presented in varying patterns. In some cases there is not only description of the apparatus itself but also a fairly detailed exposition of the experimenter's manipulations: how to smoke kymograph paper and fix the record, how the oscillograph beam can be shifted on the tube face, how scalp electrodes are applied and removed, how the visual field is mapped in perimetry. In other instances (e.g., in the consideration of timing devices), it seems more appropriate simply to describe the equipment, perhaps with the aid of a wiring diagram.

It is explicitly stated that the book is intended to be an introduction rather than an ultimate guide to research. The primary effort is to suggest by illustration the types of question that must be asked when apparatus is adopted to extend measurement and control. It is abundantly clear that the practical choice among procedural alternatives is specific to the requirements of the particular experimental situation, often in the nature of a compromise that is constantly under review as additional technical refinements become available.

The text is liberally sprinkled with line drawings and cuts, with an average of one or more every two pages. Many are helpful, a few are not at all illuminating (some exterior views of commercial items are prime examples in this category), several are confusing because of inconsistent or incomplete legends, and one or two are misleading. There is some unevenness in the inclusion of references for advanced study, but the author offers a mimeographed bibliography of more than 1,000 relevant items. Altogether, the usefulness of the book far outweighs its limitations, and the advanced student or the teacher in laboratory courses will find it of considerable help in surveying the technical tools of the trade.

FRANK W. FINGER
University of Virginia

BECK, SAMUEL J., *The six schizophrenias: reaction patterns in children and adults*. New York: American Orthopsychiatric Association, Inc., Research Monograph No. 6, 1954. Pp. viii+238. $5.00.

A coordinated inquiry into problems of schizophrenia, as evaluated and dealt with by psychiatric method, psychological method, and social work method, is the aim of this book. In a study of sixty children and fifty adults, the author attempted an interdisciplinary approach, built mainly around diagnostic Rorschach test evaluations made by psychologists and clinical behavioral evaluations made by psychiatrists, thus leading to a correlative approach between psychiatric findings and Rorschach test results.

The outline of clinical psychiatric formulations and their Rorschach correlates includes such areas as (*a*) the defenses, as assessed through obsessive-compulsive behavior, withdrawal from stimuli, projection, rigid hold on self and on reality, metaphysical and philosophical ponderings, and other defenses; (*b*) the ego and its functions, as seen in motor, perceptive, and thinking functions and total behavior; (*c*) status of emotional forces, as evaluated in degree and type of anxiety, emotions independent of external stimulus, degree of dominance of fantasy life, mood, level of organization of feelings, and (*d*) restitutional forces, as seen in religiosity, rebirth fantasies, and world reconstruction fantasies.

This, then, constitutes the conceptual tool of this study. It emphasizes a personality construct in which ego functioning, dynamically viewed, is the center of focus.

This tool resulted in a set of 120 psychiatric formulations for describing schizophrenia and their paired Rorschach indicators, each pair of which, for this research, may be regarded as a postulate. These psychiatric items were then used by the psychiatrists in describing each case, while the psychologists, working independently, made their description of each case from the Rorschach findings. By use of the Q technique (expertly described by William Stephenson in his chapter, "Q-Technique and the Rorschach Test"), these independent psychiatric and psychological appraisals were then correlated and factored for each patient. The factor analysis of the data resulted in a trait universe for six schizophrenic patterns which are summarized in Table 1 (Beck's Table 5, p. 132). The factors for the six schizophrenias, using the scheme of Table 1, are shown in Table 2 (Beck's Table 6, p. 132).

As is apparent, emphasis in this study was on a description of the manifestations and characteristics of schizophrenia, with light shed on this challenging psychological disorder from this point of view. One might wish, however, that these data had been *used* more by way of interpretation and integration for theory

TABLE 1

CONCEPTUAL SCHEME OF FACTORS IN SCHIZOPHRENIA

Theoretical Component	Factor	Level of Manifestation	
A. Defense Organization	I′	a. Withdrawal c. Pathogenic	b. Constriction
B. Intellectual Functioning	II′	a. Orderly	b. Disrupted
C. Fantasy Activity	III′	a. Autistic c. Little or none	b. Regressive
D. Social Adaptation	I″	a. Self-absorption c. Restitutional	b. Self-deprivation
E. Emotional State	IV′	a. Liability	b. Fixed Tone

TABLE 2

FACTORS FOR THE SIX SCHIZOPHRENIAS

Schizophrenia Types	Characterizing the Types				
S-1	A. c	B. b	C. c	D. a & c	
S-2	A. a & c	B. b	C. a & b	D. c	E. a
S-3	A. b	B. a	C. c	D. b	E. a & b
SR-1	A. b & c	B. b	C. c	D. b	E. a
SR-2	A. a & c	B. a	C. a & b	D. a	E. a
S-G	A. b	B. b	C. c	D. a	E. b

building concerning schizophrenic disorders, their origin, development, dynamics, and treatment. This wish is strengthened all the more because one is left with the desire for more integration of these data with other research on this problem. How earlier views of schizophrenia need to be modified as the result of this work is not sufficiently elaborated.

One wonders also what ideas these findings prompted in a Rorschach expert such as Beck concerning diagnostic tools for finer screening (detection and evaluation) of schizophrenic manifestations.

One might say that this study is more concerned, in a sense, with a comparison of the psychiatrists' and psychologists' formulations about schizophrenia than the value of these findings for a better understanding of schizophrenia. Such "profession"-orientation, while necessary as part of scientific "tool" evaluation, has its *raison d'etre* in a continued problem-orientation.

This volume will interest not only psychopathologist and therapist, but all persons interested in efforts at a rapprochement of dynamic and trait approaches to personality theory.

GEORGE F. J. LEHNER
University of California, Los Angeles

KENDALL, PATRICIA. *Conflict and mood.* Glencoe, Illinois: The Free Press, 1954. Pp. 182. $3.50.

The studies reported in this small, lithoprinted volume are devoted to identifying factors leading to shift in responses to repeated attitude survey questionnaires and interviews. Analysis of panel studies conducted by the Bureau of Applied Social Research of Columbia University had shown that where first interviews were plotted against the second, marginal totals were generally quite constant, but sizable variation occurred in the cells. This instability of response, i.e., shifting from "agree" in the first interview to "disagree" in the second or vice versa, is called turnover. A special index of turnover was derived by Lazarsfeld, based on his latent structure model.

Conflict and mood were hypothesized as prime factors specifically related to response instability. Panel study results were analyzed and college student samples were used experimentally to determine if changes in responses upon repeated administration of questionnaires were associated with questions designed to produce conflict, and with self-ratings of mood states. By and large, the hypotheses were borne out. Additionally, panel studies were analyzed to identify other factors related to response instability: degree of respondent interest and concern with topic of question, question ambiguity, amount of forcing of opinion, and degree to which estimation of facts was required of interviewer or respondent. These additional factors generally appeared to be related to instability in response.

From a methodological viewpoint, these studies are important in giving specification to factors that should be controlled in questionnaire construction. Respondent conflict and mood were the major variables studied. This reviewer however, would seriously question if one of the other factors was not indeed affecting the experimental responses analyzed in terms of conflict and mood. This other factor is the degree of respondent interest and concern. The author points out that this is a dimension of increasingly recognizable importance in surveys and that lack of interest leads to perfunctory, unstable responses. One gets the impression that this factor should have been taken into account in the college student experiment, and that instability of response may well have depended upon it to a sizable extent.

The fourfold table is typical of the sociologist's approach to data analysis. Correlation analysis, linear scaling, and the like would undoubtedly make this book more to the liking of psychologists in that a more refined degree of estimation of the influence of the variables would be possible. But this in no way invalidates the findings at the level of analysis used. One final point: in the reviewer's opinion, this report would have been better produced as a monograph with an appropriate sociological research title instead of a flashy, clinical-sounding one.

RALPH R. CANTER

The Rand Corporation
Santa Monica, California

PIÉRON, H., REUCHLIN, M., BIZE, R., BÉNASSY-CHAUFFARD, C., PACAUD, S., & RENNES, P. *Traité de psychologie appliquée.* Livre Troisième. *L'Utilisation des aptitudes. orientation et sélection professionnelles.* Paris: Presses Universitaires de France, 1954. Pp. lx+342–757.

This 415-page book is volume three in a projected seven-volume work on applied psychology being brought out under the editorship of H. Piéron. Four of the six co-authors contribute a chapter each, a fifth, two and the sixth, three chapters to the book. The nine chapters are divided into two parts, five of the chapters being subsumed under the general heading of professional orientation, and the remaining four under that of professional selection.

In the opening chapter of the work, the editor of the series calls attention to the importance of becoming professional regarding the utilization of aptitudes and the scope of his appeal is nationwide. The other four chapters in the first part deal successively with surveys regarding current knowledge of different kinds of aptitude, the criteria in terms of which prediction regarding vocational success can be made, the varying demands of different trades and occupations on human resources, and different modes and degrees of adaptation made by the individual upon entering a profession. In the last chapter of this part are also considered the special problems of fatigue, monotony, accidents, and handicaps. The longest chapter in the book, covering one hundred pages or almost one fourth of the entire work, is that concerned with the specific demands of different kinds of tasks and occupations.

The four chapters of part two are subsumed under the general heading of professional selection. Treatment in this part of the work is perhaps more specific than that in some chapters in the first part. In the first chapter here different aspects of the comparatively new province of human engineering are treated. Ways in which psychological and psychophysiological studies in a variety of tasks have led to greater economy and efficiency of effort are indicated. The various factors which have been found to contribute to error are evaluated in the second chapter. The important concept of validity is considered in relation to the use of batteries of tests in the third chapter. The final chapter in the book deals with the problem of classification in government employment, in industry, and in other spheres of action.

To each chapter is appended a quite extended bibliography, except that one bibliography is given for the three chapters contributed by the same author. Most of the references are to comparatively recent studies, very few going back beyond the time of World War I. Among the few whose works are mentioned as far back as the opening of the twentieth century and before are F. Galton, J. McK. Cattell, A. Binet, J. Jastrow, and T. L. Bolton. Excepting one list in which all but one of the references are in French, the lists include in several cases at least a majority of studies in English.

It is not yet possible to evaluate this book in relation to the rest of the series of which it is a part. In itself, the work deals with aspects of general personnel problems which have been receiving increasing attention in the past two or three decades. The cooperation of several specialists seems well justified, since it would not easily be possible for a single writer to survey the literature in the many aspects of this complex problem which has manifested such rapid development in a very few years. In spite of some differences in literary style, the treatment by different writers is uniformly high and the plan of organization is well conceived. The work perhaps would be more easily used if a general index had been added.

This work represents a survey of the variety of ways in which psychological methods—testing, factor analysis, job analysis, etc.—have been used in connection with different aspects of the complex personnel problem. Activity on this psychological front recently has been rapid. It is to be expected that in the future existing methods will continue to be adapted to the needs of still other practical problems, and that new methods will be developed to meet the needs of analytic studies regarding aspects of the problem to which the old ones do not apply. As a starting point for further efforts in this important field, this work is bound to serve a very important purpose.

MICHAEL J. ZIGLER
Wellesley College

BACH, GEORGE H. *Intensive group psychotherapy*. New York: Ronald, 1954. Pp. xi+446. $6.00.

Bach has had a lot of experience with group psychotherapy, and he writes about groups with a fine combination of scholarship and clinical sense. His book is an exciting one, for a number of reasons. First it presents a new concept of how to work with groups for therapeutic purposes. Then it richly embroiders on its central thesis through a large number of descriptions of clinical problems and their solution in the group. There is an astuteness in conception and a dispatch in execution of clinical tactics that a less bold group therapist observes with respectful admiration. Then there is wide-ranging scholarship, with citations from many fields of psychological literature, illuminating myriad aspects of the group therapy process. But finally, and I regret to have to write this, the book has annoying qualities which detract from its effectiveness. It should be a much better book than it is. Mainly it is lacking in discipline. Theory and clinical practice are not integrated; the precise relevance for intensive group psychotherapy of citations from the literature is often illusive; imprecisions, overgeneralizations, and inconsistencies are all too frequent. But let's be more specific, mostly about the strengths of the book, for these considerably outweigh its weaknesses. The weaknesses one may regard as minor annoyances that can be lived with more or less comfortably while finding out how Bach does psychotherapy with groups. There's a lot to learn about that.

The main idea of intensive group therapy is to heighten the therapeutic significance of the group by requiring that the conflicts of the individual be worked out in his immediate and contemporaneous relationships with other members of his group. The group experience is not an auxiliary to other therapies; on the contrary, all therapeutic endeavors are made to coalesce in the group. That group interactions provide opportunities for clarification of a person's difficulties and for practice of new modes of relating to others is an idea common to many approaches to group therapy. What is new and significant here, as the name of the method implies, is the *intensity* of the relationships that Bach's methods generate. How does he intensify the relationships? A careful reading of the first section of the book will provide abundant illustrations, only a few of which can be given here. For one thing, the groups meet frequently, once or twice a week for two hour periods, and for a large number of sessions (four of nine patients described had met more than 170 times). The two-hour office sessions are followed by a required two-hour postsession at dinner without the therapist being present. Postsession occurrences become appropriate material for analysis in the next formal hour. When individual sessions are held with group members, the focus is on the group experience, and the expectation is that the events of the individual sessions will be shared with the group. New participants are encouraged to record their reactions to members of the group and then read their notes to the group. Dreams, drawing, role-playing, and therapist-planned activities are used to actualize the patient's problems in the group. Bach is very inventive in finding ways of bringing a person's problems into focus in the context of the group. There is an accent on experience, on practice in relationship to people present, on deriving meaning from the immediate experience in the group: " . . . the patient is literally 'caught in the act' by his own behavior." Especially helpful in the clinical section of the book is the inclusion of verbatim protocols and reproductions of drawings produced in the groups.

Bach's collections of "do's and don't's" for group therapists is a good example of how profitably he has reflected on his experiences; these admonitions could be used as a framework for an entire course on group psychotherapy. What of outcomes? Bach is not alone among group therapists in being able to say little about this. While there is no quantitative evaluation of the effectiveness of the method, there is a consistent clinical viewpoint which has been sufficiently well defined to be tested in the future.

The book aspires to integrate clinical practice with Lewin's field concepts, but the effort does not quite come off. Though Part III of the book turns theoretical and elaborates on the implications of group dynamics for group psychotherapy, the methods recommended by Bach seem less required by the theory than one is led to anticipate. They are not in conflict; they just don't mesh in any convincing way. Bach is obviously a very creative clinician; I surmise that he also has a scholarly conscience that makes theory attractive even though not compellingly relevant. In Part II, however, are many sound and provocative observations about groups. They hardly constitute an integrated theory, but they are valuable observations—a step between clinical practice and a theory.

Now for the annoyances. Bach "writes loose," and it is often hard to know just what he believes. What is his position with respect to tests, for instance? He "always" gives an elaborate battery of tests to prospective group members and then dismisses test results as "neither specific enough nor stable enough for the prediction of actual interpersonal behavior in groups." With unabashed confidence, which seems inversely proportional to the amount of solid evidence on the subject, he dismisses the family as a psychologically significant source for experience for modern man and substitutes the peer group. He rejects, soundly I think, the family as a model for the group therapy situation. But one of his arguments—that there are too many different kinds of family patterns to use "the family" as a model—he ignores when he recommends the peer group as a model. Surely there are hundreds of different patterns of peer group relationships. Finally, to mention only one other item, Bach seems occasionally to jetison the essential concept of probability in behavior and describes events as occurring "always," "never," "inevitably," and "invariably." As I said, it's loose. But such distractions are minor if one keeps turning to pages where Bach describes his clinical procedures and his immediate rationale for them.

NICHOLAS HOBBS
Peabody College

LEWIS, HILDA. *Deprived children.* New York: Oxford Univer. Press. 1954. Pp. xviii+163. 9s. 6d. net.

This compact and carefully written book is a report of a social-clinical study of 500 children entering the Mersham Reception Centre in Kent, England, between October, 1947 and July, 1950. Three concerns are evident: (*a*) an evaluation of the work of the Centre; (*b*) an evaluation of the effects of earlier maternal separation on the child's adjustment at the time of admission and on his adjustment two years later; and (*c*) an evaluation of the factors leading to the child's placement at the Centre and the importance of these factors for his subsequent adjustment.

The writer has with care considered the family backgrounds, socioeconomic influences, and personal experiences of these children prior to their admission, and she has related these variables to the personality and behavior patterns of the children. The large number of variables which she considers in these areas and the skill with which she has interrelated these factors, sets this research report in marked contrast to the speculation characterizing most of the reports on deprived children. While this is no small achievement, perhaps her main contribution has been in showing that scientific methodology and statistical techniques may be fruitfully used in investigations in this area.

A number of limitations on the

data on which the report was based should be noted, although perhaps most of them were acknowledged by the writer; the diagnosis, ratings, and evaluations were apparently carried on by only one individual, the writer—hence, their reliabilities are indeterminable. There was no control group and there is question as to whether the compared groups were always comparable on variables other than on the one under consideration. The criteria employed were not always specified, although it must be granted that they were given more precisely and frequently than in any other study in the area to date. While there is little doubt that the foster homes were superior to the "true" homes in many respects, in what respects it is impossible to determine from the book. Such information would have added materially to our knowledge of what kind of children should be placed in what kind of homes, and to the value of an already excellent book.

The author found in her investigation that the degree of disturbance of the child entering the Centre is related to his mother's emotional instability, her intellectual level, her tendency for overindulgence, her tendency to neglect and/or to reject him, etc. However, she found that there was not a significant relationship to the presence or lack of availability of the mother. She found that "parental rejection" was significantly related to the child's manifestation of "unsocialized aggressive behavior," that "neglect and bad company" were significantly related to "socialized delinquency," and that a background of "constraint" was significantly related to "neurotic behavior." In her two-year follow-up she found that the condition of those placed as recommended was superior to those not placed as recommended and that there had been considerably greater improvement in the former group. The preceding results are only a sample of the findings she reports.

The author is to be commended for her research contribution, and for her very readable report. The book should be read by all concerned with the welfare of the child.

Samuel R. Pinneau
University of California

Arnheim, Rudolf. *Art and visual perception.* Berkeley and Los Angeles: Univer. of California Press, 1954. Pp. x+408. $10.00.

Although art is one of the most fruitful areas of application of scientific psychology, it has received relatively little attention from psychologists. One can count the books dealing with art by psychologists on the fingers of one hand, and psychological studies dealing with art are few and far between. In reading this book, one realizes why more psychologists have not been concerned with art. Art is a technical specialty in its own right and one must be expert both in psychology and in either creative art or the history of art to write on art. Arnheim's book brings the scientific knowledge of a trained psychologist to bear on the fundamental problems of visual art as it has developed through the ages. The discussion is always with reference to concrete works of art. Paintings, designs, figures, and sculptures are chosen to present the problems of balance, shape, form, growth, space, light, color, tension, and expression operationally, almost experimentally. Many original drawings, diagrams, and figures illustrate basic principles and important points. The writing is superb. The book is full of penetrating insights into questions of art and also into many problems of concern to the psychologist.

Fundamentally this book is an argument against the usual art historian's approach, so well described by Arnheim as the purely subjective point of view, that what a person sees in a work of art (and, we may add, what the artist creates) "depends entirely on who he is, what he is interested in, what he has experienced in the past, and how he chooses to direct his attention" (p. 64). Against this view, Arnheim marshals visual phenomena independent of individual experience: phenomena demonstrating simplicity, subdivision of wholes, grouping, formed patterns, and the other many properties well known to psychologists through the work of the Gestalt group. Three basic principles, of which one is methodological and two are explanatory, dominate the author's argument. The methodological principle assumes that immediate perception furnishes the clues to the work of art, since it reveals the figural properties determining form orientation, tridimensionality, subdivision, and grouping of complex patterns. If "these patterns have an objective structure of their own, they are likely to offer a solid basis on which the artist can rely" (p. 64). The two explanatory principles are simplicity of structure and dynamics of brain processes. The importance of these principles can perhaps best be appreciated by noting that they serve in place of learning, culture, historical accident, and individual genius to account for why things look as they do and why artists draw and paint as they do.

The degree to which the reader will accept Arnheim's principal assumptions is probably a personal matter. This reviewer goes along with the evidence of phenomenology, immediate experience, up to a certain point, and then finds he cannot go along the rest of the way. Thus, he is willing to accept phenomenological evidence that what we accept as good representations of reality is a matter of familiarity and adaptation, and changes with the times. We are unaware of the distortions of our own time and culture. But he cannot accept the author's explanation of children's drawings and paintings as the result of the fact that they reproduce what *they* see. This account is not convincing to an adult such as the reviewer, whose drawing ability never went beyond that of the lower quartile of six-year-olds. In view of the inchoate character of much modern art, it may be naive to assert that the artist must learn how to express what he sees and what he wants to convey just as the child must learn a language to express what he wants to say. And even though it may be granted that modern art requires a technique as much as representational art, nevertheless, *it* must be learned, although it is a different technique from classical representational art.

Similarly, one may follow the author up to a certain point in his reliance on the principle of simplicity to account for the way things appear as they do. For the fairly simple figures considered by Arnheim we may agree with him that ". . . when three-dimensional perception occurs (with certain figures) we must assume that undistorted shape in a tilted position makes for a simpler total situation than distorted shape in frontal position" (p. 205). But we see many distorted shapes in the frontal-parallel plane which do not go three-dimensional to be seen more simply; and as I write, my desk beside me is a wirr-warr of regular shapes in random orientations giving an impression of disorder which no simplifying tendency of perception, no balancing of brain forces, makes orderly. And so with much in art, particularly modern art.

Arnheim recognizes that simplicity may be overworked for " . . . if the tendency to simplicity persisted all the way, there would be no seeing whatsoever" (p. 359). In his admission that the reality-level of art changes with familiarity he recognizes that learning and cultural factors also operate to determine what is perceived. That the traditional use of learning and familiarity factors in *explaining* perception was inadequate few would now deny, but this fact should not prevent us from incorporating their true features into more adequate theories. At this point, the reviewer makes bold to assert, adaptation-level theory could have been of service in providing a rational basis for the operation of learning factors in perception.

The discussion of color leaves one who has devoted considerable attention to this subject somewhat confused. Present-day psychologists do not speak of green as yellowness plus blueness, even though they know that blue and yellow pigments yield green when mixed. Nor is orange the complement of blue. An arrangement of the hues in a color circle that does not include green as a "fundamental" color (primary, unitary?) will also puzzle psychologists. In denying that complementary colors can be specified in terms of secondary standards (pigmented papers, for example) Arnheim overlooks the fact that both the Munsell and Ostwald color systems specify complementaries as well as color matches. Nor does Arnheim make use of the best data on color preferences in his theory of color harmony. While the experimental work on color harmony is not fully adequate for explaining all the uses of color by artists, psychologists would be happier if Arnheim had made more use of such experimental results as are to be found in the literature (Guilford's, for instance) in presenting his own theory of color harmony.

These criticisms are offered in a constructive spirit, for no one reading this magnificent book can fail to appreciate its splendid contributions. It has illumined many purely psychological problems for the reviewer as well as widened and deepened his appreciation of art. And a book which reflects so well the author's urbanity, catholicity, and keenness of mind, as well as his technical grasp of the scientific and the artistic, is no small achievement. It should be in every psychologist's library and a required reference in courses dealing with vision and perception.

HARRY HELSON
University of Texas

MEEHL, PAUL E. *Clinical versus statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. x+149. $3.00.

Meehl attempts to plot a course to the port of valid prediction through the "rigorous" channels of statistical methods and the "sophisticated" undercurrents of clinical dynamics. If these adjectives seem prejudicial, one may select others from four lists of some 20 words each (p. 4), all of which are very handy for stereotyping either the statistical or clinical methods of prediction. Terms like operational, communicable or verifiable may be used when statistical methods are favored while mechanical, atomistic, or additive are available if such methods are disliked. Similarly, proponents of clinical methods may favor holistic, patterned, or organized, while opponents may prefer mystical, subjective, or intuitive. As the author points out in the preface, students have reacted to his lectures on this topic as to a projective technique and it is expected that readers will do likewise according to their biases.

Most of the book is devoted to a thoughtful but discursive analysis of the alternative strategies for prediction of statistical or actuarial methods as compared with clinical or case-study methods. Meehl believes that much confusion could be avoided if careful distinction were made between two different uses of statistics, the types of data involved in prediction, and the methods of combining data in making predictions. One use of statistics is referred to as discriminative or validating while the other is called structural or analytic. Although this differentiation is debatable, there can be little quarrel with the conclusion that predictions, however arrived at, must ultimately be evaluated by statistical procedures —"always . . . the shadow of the statistician hovers in the background" (p. 138). In contrast Meehl thinks there is little likelihood that clinical hypotheses can ever be generated by formal rules of logic and statistical operations.

The paucity of empirical evidence regarding the relative "batting averages" of clinical vs. actuarial prediction is emphasized in Meehl's summary of some 20 possibly pertinent studies. By some rather heroic analysis, Meehl concludes that "in all but one . . . the predictions made actuarially were either approximately equal or superior to those made by a clinician" (p. 119). Unhappily, it is questionable how closely any of the studies approaches an ideal comparative design. There appears to be a clear need for more exacting and systematic comparison of the two methods of prediction with respect to both predictive efficiency and relative costs. Meehl does an excellent job of raising some of the basic problems and issues which have to be considered in carrying out research in this area.

LEONARD S. KOGAN
Institute of Welfare Research
Community Service Society of New York

ROTTER, JULIAN B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954. Pp. xvi+466. $6.50.

In the preface Rotter says: "The purpose of this book is to arrive at a systematic theory from which may be drawn specific principles for actual clinical practice, and to illustrate some of the more important applications of the theory to the practice. Rather than attempt to apply this theory to all the problems facing the clinical psychologists, we have chosen to apply it to only two of the clinician's most important problems—the measurement of personality (personality diagnosis) and psychotherapy. Even in these broad areas the application is more illustrative than comprehensive."

The first three chapters ("The Importance of Theory in Clinical Psychology," "Some Major Problems of Clinical Psychology," "Criteria for a Language of a Description for Clinical Psychology") represent for the most part a clear and incisive introduction to the major purpose of the book, chapters which can be read with profit by all clinical psychologists. The next four chapters, which represent the bulk of the book, contain the aims and concepts of Rotter's social learning position as well as the ways in which it differs from other approaches. It is impossible in the scope of this review to describe and evaluate the basic concepts of Rotter's position. There is, however, one concept which deserves, in this reviewer's opinion, special emphasis, namely, expectancy: the " . . . probability or contingency held by the subject that any specific reinforcement or group of reinforcements will occur in any given situation or situations." Although, as Rotter points out, this is not a new concept, he gives it an importance which it has heretofore not had in systematic theory. It would perhaps be more correct to say that he states this concept in a way which has great relevance for clinical practice, although he does not pursue this relevance as much as this reviewer would have liked. Few, if any, would quarrel with Rotter when he says: "The understanding and prediction of changes in expectancies is one of the more crucial aspects of this theoretical approach. Not only is it true from our formulations that behavior depends, in part at least on expectations, but reinforcement values themselves depend upon expectations for subsequent reinforcements. Consequently, understanding how expectations change becomes of crucial significance for any application of this theory to psychotherapy or any effecting of behavioral change." Unfortunately, in the chapter on psychotherapy one does not find examples or descriptions of how expectancies are deduced by the therapist, how they change as therapy progresses, and how awareness of the importance of expectancies determines therapeutic tactics and outcomes. To do justice to what Rotter considers to be crucial problems it would have been necessary to present and discuss the data of psychotherapy. It is more than surprising that not a single therapeutic interaction, or any part of one, is presented. In terms of the stated purposes of this book, this lack is indeed serious. In fact, in the entire book there is not anything (with one possible exception) resembling a clinical case—a lack which makes one wonder why "clinical psychology" is in the title of the book. Is it carping or hypercritical to expect that an attempt to apply a systematic theory to clinical problems would involve presentation and discussion of the data with which the clinician deals? Can you talk very meaningfully about psychotherapy without therapeutic data?

The above criticisms, stated in terms of the concept of expectancy, hold for most of Rotter's other formulations. To make the objection clear: this reviewer feels that Rotter has not come to grips with those clinical problems which his theoretical formulations were intended to illuminate. If he had given as much systematic thought to the clinical problems as he did to the statement and buttressing of his theoretical formulations, this book would have been more of a significant contribution than it is.

There is one other major criticism which has to be made. Rotter's discussion and evaluation of psychoanalytic theory is amazingly superficial and, to the unwary graduate student, misleading. For example: "Questions such as why children around the age of six show considerable attachment to parents of the opposite sex, why people become aggressive following frustration, why people show attraction to the same sex, and so on, are answered [by psychoanalysts] 'They have inherited an energy that directs them to do it.' Freud did state that the particular form which the expression of the energy took was perhaps a function of learning or experience and that behavior itself, over a long period of time or in adulthood, was determined by what happened to the person when these instinctual energies were expressed in early childhood. To the ultimate 'whys' Freud could say little other than because it is an instinct." This reviewer submits that this quotation, which is fairly representative of Rotter's handling of Freud, reflects a distorted and biased understanding of Freud's position. But why does Rotter focus only on Freud?; he explicitly points out that there have been changes in psychoanalytic theory. An author has to select what he will talk about—but is age of the concept an adequate criterion for selection? Should an evaluation of learning be concerned only with Pavlov?

Some very positive things should be said about this book. First, it represents one of the few attempts to formulate and apply a learning theory to clinical phenomena and problems—the more such courageous attempts we have, the better will we be able to evaluate the adequacy of such theories. Second, Rotter's formulations have generated a relatively large number of studies at The Ohio State University—a tribute not only to Rotter's effectiveness as a teacher but a reflection of the fruitfulness of the formulations. As Rotter and his students continue to apply these formulations to complex clinical phenomena, we will be in a better position to evaluate the degree of fruitfulness of their theoretical approach.

Seymour B. Sarason
Yale University

Hyman, H. H., *et al. Interviewing in social research.* Chicago: Univer. of Chicago Press, 1954. Pp. xvi+415. $8.00.

This volume is a scholarly but free-flowing treatise on sources of error and potentialities for their control in the *survey interview.* It is the report of a six-year collaborative undertaking by the National Opinion Research Center (NORC), commissioned by the Joint Committee of the Social Science Research Council and the National Research Council on the Measurement of Opinion, Attitudes, and Consumer Wants, and sponsored by the Rockefeller Foundation. A workmanlike job has been done of rendering a pattern of reasonable speculations from a conflictful mass of evidence and pseudo evidence. Although results would have appeared more consistent had the authors confined themselves to studies carried out under their own auspices, they elected to attempt the more difficult but valuable task of integrating their own findings with the available literature, studies by other agencies, and the opinions of experienced workers in the field. The end product is an excellent and up-to-date summary of an obviously complex subject.

The book is divided into seven rather lengthy chapters. Chapter I, entitled "A Frame of Reference for the Study of Interviewer Effects," reviews the widespread application of interviewing as a fallible method of scientific inquiry, and cautiously suggests the possibilities for generalization about other kinds of interviewing from studies of the survey interview. Interview error is noted as being of two major types—inter-interviewer variation or reliability and accuracy or validity. Somewhat surprisingly, the authors appear to entertain the notion that maximum reliability is somehow incompatible with maximum validity, each being a function of differential locus along a continuum of freedom allowed the interviewer.

Chapter II, "The Definition of the Interview Situation," is devoted to the formulation of hypotheses about factors affecting interviewing from "phenomenological" case studies based largely on intensive interviews with interviewers and respondents.

The procedures used in these case studies are elaborated in Appendix A. Some will note with satisfaction the special role of the nonsurvey interview in developing a framework for inquiry about the survey interview.

The third chapter presents an analysis of "Sources of Effect Deriving from the Interviewer." The major conclusion drawn from a series of studies is that the interviewer's own opinions (ideological bias) generally affect the results of survey interviews to a less important degree than do his expectations based on the respondent's early answers (attitude-structure expectation), judgment of the respondent's group membership (role expectation) or the interviewer's beliefs regarding prevailing opinion in the population (probability expectation).

Chapter IV, "Respondent Reaction in the Interview Situation," reviews the contradictory evidence on personal interviewing vs. self-administered questionnaires as well as the systematic and differential effects of disparities between interviewer and respondent with regard to sex, age, group membership, class, and so on. The response of the respondent in the interview is seen to be a function of his task involvement and his social involvement.

The "Situational Determinants of Interviewer Effects" are examined in Chapter V. Among such determinants which influence the psychological processes of the interviewer and respondent, and their interaction in the interview, are respondent anonymity, survey sponsorship, structuring of the interview procedure, form and content of the survey, and method of recording.

Chapter VI is an attempt to evaluate the magnitude of "Interviewer Effects under Normal Operating Conditions." On the basis of an admittedly small body of evidence, the conclusion is reached that "for most fixed-response opinion questions there is relatively little inter-interviewer variation." Although this conclusion seems somewhat optimistic in the light of the evidence, it seems clear that interviewer variation tends to be relatively smaller for fixed-response opinion questions and factual questions than for free-response questions and field ratings by interviewers.

The final chapter is devoted to consideration of the relative merits of various methods for "Reduction and Control of Error." The first portion of the chapter implements the findings presented in preceding chapters by describing possibilities for control of error arising from the interviewer through better selection and training, control of error arising from respondent reactions by matching characteristics of interviewer and respondent, and control of situational errors by standardization of presumably optimal interview procedures. The final section treats of the indirect control of error by statistical estimation of its magnitude and components.

In recognition that many of the findings reported are based on and may be limited to the kinds of interviewers hired by NORC, Appendix B describes NORC training and field procedures, and the characteristics of the NORC national field staff.

LEONARD S. KOGAN
Institute of Welfare Research
Community Service Society of New York

ERRATUM

In the review of Roger G. Barker of *Manual of Child Psychology*, edited by Leonard Carmichael, which appeared in the May 1955 *Psychological Bulletin*, there is a typographical error. The sentence on page 266, right-hand column, should read "However, this reader found intriguing the world view so *boldly* sketched."

INDEX OF AUTHORS

ARTICLES

Ammons, R. B., 69
Auld, Frank, Jr., 377
Barnette, W. Leslie, 105
Bevan, William, 473
Birney, Robert, 51
Brayfield, Arthur H., 396
Bruner, Jerome S., 122, 376
Crockett, Walter H., 396
Cronbach, Lee J., 177, 281
Dukes, William F., 24
Ellis, Douglas S., 425
Emmons, William H., 328
Farber, I. E., 311
Fiske, Donald W., 217
Kogan, Nathan, 122, 376
Livson, Norman H., 77
Lord, Frederic M., 505
McHugh, Richard B., 425
McKeachie, W. J., 51
Meehl, Paul E., 194, 281
Melikian, Levon H., 303
Murray, Edward J., 377
Nichols, Thomas F., 77
Pinneau, Samuel R., 429, 459
Pressey, S. L., 343
Prothro, E. Terry, 303
Ray, Wilbert S., 134
Rice, Laura, 217
Rosen, Albert, 194
Sato, Koji, 345
Simon, Charles W., 328
Spitz, René A., 453
Taft, Ronald, 1
Tagiuri, Renato, 122, 376
Willerman, Ben, 132

BOOK REVIEWED

Ainsworth, M. D., 360
Allport, Floyd H., 521
Anastasi, Anne, 97
Arbuckle, D. S., 251
Arnheim, Rudolf, 537
Arnold, Magda B., 88, 372
Arsenian, S., 251
Atkinson, J. W., 95
Ausubel, David P., 100
Bach, George H., 534
Bachrach, Arthur J., 358
Baker, Harry J., 150
Beck, Samuel J., 530
Bellak, Leopold, 370
Bellows, Roger M., 98
Bénassy-Chauffard, C., 533
Berg, Irwin A., 276
Bize, R., 533
Blair, Glenn M., 511
Braatøy, Trygve, 171
Brand, Howard, 465
Bray, Douglas W., 172
Brosin, H. W., 361
Bross, Irwin D. J., 101
Brown, Judson, S., 94
Burt, Cyril, 469
Carmichael, Leonard, 263
Carroll, John B., 87
Cartwright, D., 361
Chapman, A. L., 251
Christie, Richard, 83
Clark, R. A., 95
Commins, W. D., 511
Coombs, C. H., 267
Cooper, Linn F., 356
Craig, R. C., 251
Cronbach, Lee J., 511
Crow, Alice, 150
Crow, Lester D., 150
Cruze, Wendell W., 150
Davidson, Audrey, 150
Davis, R. L., 267
Driver, Helen Irene, 251
Dubin, Robert, 270
Dupertuis, C. Wesley, 367
Dymond, Rosalind F., 528
Edwards, Allen L., 470
Eldersveld, S., 361
Emery, F. E., 468
Erickson, Milton H., 356
Estep, M. Francis, 98
Estes, William K., 161
Fagin, Barry, 511
Fairbairn, W. Ronald D., 366
Fay, Judith, 150
Fink, Harold K., 98
Fitts, P. M., 361
Frick, F. C., 361
Fruchter, Benjamin, 99
Gannon, Timothy J., 169
Ganz, Madelaine, 150
Gasson, J. A., 88
Gillin, John, 165
Ginsburg, Sol W., 172
Ginzberg, Eli, 172
Good, Carter V., 463
Grant, D. A., 361
Gray, J. Stanley, 85
Griffiths, W., 251
Grings, William W., 529
Gruenberg, Sidonie M., 274
Hall, Calvin S., 366
Halpern, Florence, 150
Hammond, S. B., 468
Harms, Ernest, 90
Hathaway, S. R., 251
Henry, Andrew F., 368
Herma, John L., 172
Hertzman, M., 79
Hoch, Paul H., 354
Hoijer, Harry, 359
Holt, R. R., 360
Hoyles, J. A., 251
Humphreys, J. A., 251
Hurlock, Elizabeth B., 150
Hyman, H. H., 542
Ingham, H. V., 174
Jahoda, Marie, 83
Jersild, Arthur T., 150
Jones, M. R., 159
Jones, R. Stewart, 511
Katz, B., 251
Katz, D., 361
Kendall, Patricia, 532
Kerman, Edward F., 466
King, H. E., 277
Klopfer, B., 360
Klopfer, W. G., 360
Knapp, R. H., 251
Kornhauser, Arthur, 270
Kugelmass, I. Newton, 470
Lawrence, M., 363
Le Beau, J., 471
Lee, A. McC., 361
Lehner, G. F. J., 251
Lewis, H. B., 79
Lewis, Hilda, 536
Lindgren, H. C., 251
Lindzey, G., 346
Little, W., 251
Love, L. R., 174
Lowell, E. L., 95
Lowenfeld, Margaret, 369
McClelland, D. C., 95
McCulloch, W. S., 361
McDermott, Eugene, 367
McKenzie, F. W., 251
McKeon, Richard, 357
McMillan, B., 361
McNemar, Quinn, 518, 525
Machover, K., 79
Martin, William E., 150
Maslow, A. H., 159

Meehl, Paul E., 539
Meissner, P. B., 79
Miller, G. A., 361
Monachesi, E. D., 251
Morgan, Clellen L., 511
Moses, Paul J., 268
Moustakas, Clark E., 150
Mursell, James L., 511
Murphy, G., 358
Nuttin, Joseph, 269
Oeser, O. A., 468
Ort, R. S., 251
Pacaud, S., 533
Peck, Leigh, 150
Pennington, L. A., 276
Pepinsky, H. B., 251
Pepinsky, Pauline Nichols, 251
Perry, Ralph Barton, 93
Piaget, Jean, 526
Piéron, H., 533
Recktenwald, L. N., 251
Remmers, H. H., 511
Rennes, P., 533
Reuchlin, M., 533
Révész, G., 523
Roback, A. A., 355
Rogers, Carl R., 528
Ross, Arthur M., 270
Rotter, Julian B., 540
Ryan, Thomas, 466
Ryden, Einar R., 511
Sanderson, H., 251
Sappenfield, Bert R., 173
Sarason, Seymour B., 168
Savage, Leonard J., 354
Scates, Douglas E., 463
Schafer, Roy, 271
Schmuller, Allen M., 273
Segal, J., 91
Shaw, F. J., 251
Sheldon, William H., 367
Short, James F., Jr., 368
Simpson, Ray H., 511
Smith, Henry P., 511
Smith, Patricia, 466
Solberg, Patricia Anne, 150
Sonnemann, Ulrich, 275
Sorenson, Herbert, 511
Special Devices Center, Office of Naval Research, 524
Stendler, Celia Burns, 150
Stolz, Lois Meek, 150
Stone, Calvin P., 518
Taylor, W. S., 272
Thelen, Herbert A., 278
Thorpe, Louis P., 273
Thrall, R. M., 267
Tinbergen, Niko, 163
Traxler, A. E., 251
Tyler, Leona E., 251
Wapner, S., 79
Warters, Jane, 251
Wever, E. G., 363
Witkin, H. A., 79
Wolberg, Lewis R., 167
Wolffheim, Nelly, 150
Wolstein, Benjamin, 363
Zilboorg, G., 362
Ziskind, Eugene, 357
Zubek, John P., 150
Zubin, Joseph, 354

BOOK REVIEWERS

Adler, Dan L., 366, 468
Albee, George W., 360
Anderson, John E., 526
Arnold, Magda B., 372
Barker, Roger G., 263, 544
Barmack, Joseph E., 357
Bayley, Nancy, 367
Beach, Frank A., 163
Blum, Milton L., 466
Brown, Roger W., 359
Canter, Ralph R., 532
Carmichael, Leonard, 172
Chapanis, A., 361
Cofer, Charles N., 272
Crissy, W. J. E., 85
Cronbach, Lee J., 267
Dashiell, John F., 521
Drever, James, 518
Ellis, Albert, 275
Ely, Jerome H., 270
Eysenck, H. J., 471
Finger, Frank W., 530
Goodstein, Leonard D., 168
Grant, David A., 101
Guilford, J. P., 97
Hall, Calvin S., 355
Hamlin, Roy M., 83, 171, 268, 271, 363
Harris, J. Donald, 363
Heiser, Karl F., 470
Helson, Harry, 537
Hilgard, Ernest R., 161
Hirsh, Ira J., 87
Hobbs, Nicholas, 276, 534
Hutt, Max L., 370
Katz, Barney, 100
Katz, Daniel, 346
Kellogg, W. N., 98
Kluckhohn, Clyde, 357
Kogan, Leonard S., 99, 274, 354, 470, 525, 539, 542
Krugman, Morris, 90, 369
Lee, Alfred McClung, 278
Lehner, George F. J., 530
London, Ivan D., 361
McCandless, Boyd R., 150
McClelland, David C., 159
McGehee, William, 98
McKeachie, W. J., 165
McKinney, Fred, 358, 362
Marx, Melvin H., 93
Misiak, Henryk, 169
Norman, Ralph D., 173
Page, James D., 277, 354, 466
Pattie, Frank A., 523
Pinneau, Samuel R., 536
Postman, Leo, 79
Preston, Malcolm G., 463
Raimy, Victor, 528
Rotter, Julian B., 167
Saper, Bernard, 94
Sappenfield, Bert R., 88, 465
Sarason, Seymour, 174, 469, 540
Shoben, Edward Joseph, Jr., 251
Simpson, George, 368
Small, Arnold M., 524
Snygg, Donald, 511
Stolurow, Lawrence M., 273
Weitzenhoffer, André M., 356
White, Robert W., 95
Zigler, Michael J., 91, 269, 533

BUREAU OF EDL. & PSYCH. RESEARCH : : LIBRARY
15 K 26
2711